

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:02:22 ; Search time 13.08 seconds
(without alignments)
679,734 Million cell updates/sec

Title: US-09-886-400-4

Perfect score: 1877
Sequence: 1 LRALVFHGNLOYAIPKSEI.....RRLDAFRAIYNDRGNGEP 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/6C-COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/6D-COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1580	84.2	346	2	US-08-613-220B-4
2	141.5	7.5	647	1	US-07-894-212A-8
3	141.5	7.5	649	1	US-07-894-212A-2
4	141.5	7.5	650	1	US-07-893-928A-1
5	104	5.5	329	1	US-08-270-013B-2
6	104	5.5	329	1	US-08-838-418-2
7	88.5	4.7	2089	1	US-08-418-893D-23
8	88.5	4.7	2089	1	US-08-418-893D-24
9	87	4.6	235	2	US-09-141-135-2
10	87	4.6	657	4	US-09-355-166-1
11	86.5	4.6	889	1	US-08-248-021A-2
12	85.5	4.6	764	4	US-09-235-451-36
13	85.5	4.6	788	2	US-08-907-166-6
14	83.5	4.4	641	1	US-07-718-535-3
15	83.5	4.4	641	1	US-08-161-999-3
16	83.5	4.4	1676	4	US-08-487-283A-2
17	83	4.4	506	2	US-08-849-480A-5
18	83	4.4	636	4	US-09-564-805-237
19	82.5	4.4	549	2	US-08-676-279-59
20	82	4.4	303	3	US-09-002-298-1
21	82	4.4	623	4	US-09-104-068-4
22	82	4.4	637	4	US-09-104-068-2
23	81.5	4.3	491	1	US-08-489-733-3
24	81.5	4.3	491	2	US-08-993-581B-3
25	81.5	4.3	806	1	US-08-451-715A-6
26	81	4.3	1065	4	US-09-412-545-2
27	80.5	4.3	550	1	US-08-484-493-2

28	80.5	4.3	550	1	US-08-484-494-2	Sequence 2, Appli
29	80.5	4.3	550	2	US-08-345-212-2	Sequence 2, Appli
30	80.5	4.3	550	4	US-09-249-003-2	Sequence 2, Appli
31	80.5	4.3	554	3	US-08-904-871-1	Sequence 1, Appli
32	80	4.3	855	2	US-08-468-558-2	Sequence 2, Appli
33	80	4.3	855	4	US-08-676-444-2	Sequence 2, Appli
34	80	4.3	1110	1	US-08-118-441-29	Sequence 29, Appli
35	80	4.3	1110	3	US-08-338-579A-29	Sequence 29, Appli
36	80	4.3	1110	5	PCT-US94-09851-29	Sequence 29, Appli
37	79.5	4.2	831	1	US-08-073-384C-5	Sequence 5, Appli
38	79.5	4.2	831	1	US-08-254-359A-5	Sequence 5, Appli
39	79.5	4.2	831	1	US-08-483-043-5	Sequence 5, Appli
40	79.5	4.2	831	1	US-08-481-238-5	Sequence 5, Appli
41	79.5	4.2	831	2	US-08-471-066B-5	Sequence 5, Appli
42	79.5	4.2	831	2	US-08-484-956-5	Sequence 5, Appli
43	79.5	4.2	831	2	US-08-757-653-5	Sequence 5, Appli
44	79.5	4.2	831	2	US-08-599-491-5	Sequence 5, Appli
45	79.5	4.2	831	2	US-08-756-386-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-613-220B-4
; Sequence 4, Application US/08613220B
; Patent No. 5958751
; GENERAL INFORMATION:
; APPLICANT: Murphy, Dennis
; APPLICANT: Reid, John
; TITLE OF INVENTION: ALPHA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,220B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-68-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-613-220B-4

Query Match 84.2% Score 1580; DB 2; Length 346;
Best Local Similarity 94.5%; Pred. No. 6.7e-159;
Matches 344; Conservative 1; Mismatches 1; Indels 18; Gaps 18;

Qy 1 LRALVFGNLOQYAEIPKSEIPKVIKAYIPVIETLIKKEIPGLNITGYTLKFLPKDIID 60
Db 1 LRALVFGNLOQYAEIPKSE-PKVIKAYIPVIETLIKKEE-PGLNITGYTLKFLPKDIID- 57
Qy 61 LVKGGIASDIEIGTSYTHAILPLPLSRVQAQVORDREVKEELFELSPKGFUPELAY 120
Db 58 LVKGGIASDIEIGTSYTHAILPLPLSRVQAQVORDREVKEELFELSPKGFUPELAY 114
Qy 121 DPTIIPALIKONGYEYLFADGEAMLSAHLNSAIKPKIPLYPHLIKAQREKPRYISYLLG 180
Db 115 DPTIIPALIKONGYEYLFAD-EAMLSAHLNSAIKPKIPLYPHLIKAQREKPRYISYLL- 171
Qy 181 LREURKAIVFEGKVTLKAVKIDIEAVPVVAVNTAVMLGIGRLPLMNPVKVSWIEBCK 240
Db 172 LRELURKAIVFEGKVTLK-VKIDIEAVPVVAVNTAVML-IGRLPLMNPVKVSWIEBCK- 228
Qy 241 NILLYGTDIEFIGYRDIAGYRMSVEGLLEVIDELNSELCLPSELKHSRGRELYLRTSSNA 300
Db 229 NILLYGTDIEFIGYRDIAG-RMSVEGLLEVIDELNSELCLPSELKHSRGRELYLRTSSNA- 285
Qy 301 DKSURIWREDEGNARLNMVYNNRGEALALAEASDARGWEPLPERRLDAPRAIYNDRGE 360
Db 286 DKSURIWREDEGNARLNMVYNNRGEALALAEASDARGWEPLPERRLDAPRAIYNDRGE- 342
Qy 361 NGEP 364
Db 343 NGEP 346

RESULT 2
US-07-894-212A-8
; Sequence 8, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-894-212A-8

Query Match 7.5%; Score 141.5; DB 1; Length 647;
Best Local Similarity 22.8%; Pred No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
Qy 25 EKAYIPVIETLIKKEIP---FGLNITGYTLKELP---KDIIDLKVGKGIASDLIELIGTSY 78
Db 27 EKCYWPFLETL---EYFNPMKVAIHTSGPLIEWLQDNRPDIIDLLSLVKRGQVEITWVAGE 84
Qy 79 THAILPLPLSRVQAQVORDREVKE---ELFELSPKGFUPELAYDPIIPAILKONGYEYL 136
Db 85 YEPVLASIP---KEDIEQIRLMKWEAKSIGDARGVWLTERVWQPELVKTLKESGIDYV 141
Qy 137 FADGEAMLSAHLNSAIKPKIPLY-PHLIKAQRE-----KRFYISYLLGLRELKKA 187
Db 142 IVD-----DYHMSAGLSKEELYWNPYTTEDGGEVIAFPIDEKLR---YLIPFRPVOKV 192
Qy 188 IKL---VFEKGVTLKAV---KDIEAVPVVAVNTAVMLGIGRLPLMNPVKVSWI----- 236
Db 193 LEVLSLIDGDESKAVFHDGGEKFGIWFGTVEWY-----EKGLREFPDR 239
Qy 237 ---BDKNILLYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
Db 240 ISSDEKINMLYTEILEKYKPRGLVYLPISY-----FEM-----SEWSLPKQAKRLF 287
Qy 282 ---SELKHSG---RELILTSSWAPDKSLRIWREDEGN---ARLNMVYNNRGEALAL 330
Db 288 VEFVNEKLVKGIKFEYKRVFVGGI---KNF-FYKTPESYNNHMKMLVSKLVNRN----- 339
Qy 331 AENSDARGWEPLPERRLDAPRAIYNDRG 359
Db 340 -----PEARKYLLRAQCNDAYWHG 358

RESULT 3
US-07-894-212A-2
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 648 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-894-212A-2

Query Match 7.58; Score 141.5; DB 1; Length 649;
Best Local Similarity 22.8%; Pred. No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLP---KDIIDLKVGKGIASDLIEIIGTSY 78
DB 29 EKYWPFLETL--EYPMKVAIHTSGPLIEWLQDNRPVIDLLRSVLKRGQVEIVVAGF 86
QY 79 THAILPLPLSRVAQVORDREVKE--ELFELSPKGFWLPELAYDPIIPAILKDNQVEYL 136
DB 87 YEPVLASIP---KEDRIEQIRLMKEMAKSIGFDARGVWLTERVWQPELVKTLKESGIDIV 143
QY 137 FADGEAMLFSAHLSAIPKIPLY-PHLIKAQRE-----KRFYISYLLGLRELKRA 187
DB 144 IVD-----DYHFSAGLSKBEELWPTYTEDGGEVIAVFPIDEKLR---YLIFRPVDKV 194
QY 188 IKL---VPEGKVTLKAV--KDIEAVPVAVNTAVMLGIGLPLMNPKKVASWI----- 236
DB 195 LEYLSLIDGDESKVAVPHDDGEGFGWPGTYEWY-----EKGWLRFFDR 241
QY 237 ---EKKDNLILYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
DB 242 ISSDEKINLMLEYLEKYKPGVLPIASY-----FEM-----SEWSLPAKQARLF 289
QY 282 ----SELKHSQ-----RELYLTSSWAPDKSLRWREDEGN---ARLNMLSYNNRGEALLL 330
DB 290 VEFVNELVKGFIFERYVFRGGIW---KNF-FYKPESNYMHKRMMLVSKLVNRRN---- 341
QY 331 AENSARGWEPLPERRLDAFRAIYND--WRG 359
DB 342 -----PEARKYLLRAQCNDAYWHG 360

RESULT 4
US-07-893-928A-1
; Sequence 1, Application US/07893928A
; Patent No. 5578479
; GENERAL INFORMATION:
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Tape
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/07/893.928A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95470/C-1197
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-893-928A-1

Query Match 7.58; Score 141.5; DB 1; Length 650;
Best Local Similarity 22.8%; Pred. No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLP---KDIIDLKVGKGIASDLIEIIGTSY 78
DB 29 EKYWPFLETL--EYPMKVAIHTSGPLIEWLQDNRPVIDLLRSVLKRGQVEIVVAGF 86
QY 79 THAILPLPLSRVAQVORDREVKE--ELFELSPKGFWLPELAYDPIIPAILKDNQVEYL 136
DB 87 YEPVLASIP---KEDRIEQIRLMKEMAKSIGFDARGVWLTERVWQPELVKTLKESGIDIV 143
QY 137 FADGEAMLFSAHLSAIPKIPLY-PHLIKAQRE-----KRFYISYLLGLRELKRA 187
DB 144 IVD-----DYHFSAGLSKBEELWPTYTEDGGEVIAVFPIDEKLR---YLIFRPVDKV 194
QY 188 IKL---VPEGKVTLKAV--KDIEAVPVAVNTAVMLGIGLPLMNPKKVASWI----- 236
DB 195 LEYLSLIDGDESKVAVPHDDGEGFGWPGTYEWY-----EKGWLRFFDR 241
QY 237 ---EKKDNLILYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
DB 242 ISSDEKINLMLEYLEKYKPGVLPIASY-----FEM-----SEWSLPAKQARLF 289
QY 282 ----SELKHSQ-----RELYLTSSWAPDKSLRWREDEGN---ARLNMLSYNNRGEALLL 330
DB 290 VEFVNELVKGFIFERYVFRGGIW---KNF-FYKPESNYMHKRMMLVSKLVNRRN---- 341
QY 331 AENSARGWEPLPERRLDAFRAIYND--WRG 359
DB 342 -----PEARKYLLRAQCNDAYWHG 360

RESULT 5
US-08-270-013B-2
; Sequence 2, Application US/08270013B
; Patent No. 5686294
; GENERAL INFORMATION:
; APPLICANT: Soyabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voigt & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 61601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/270.013B
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 164701/1993
; FILING DATE: 02-JUL-1993
; ATTORNEY/AGENT INFORMATION:

```
; NAME: Green, Robert F.
; REGISTRATION NUMBER: 27555
; REFERENCE/DOCKET NUMBER: 62321
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)3533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-270-0138-2

Query Match      5.58; Score 104; DB 1; Length 329;
Best Local Similarity 26.08; Pred. No. 0.011;
Matches 40; Conservative 24; Mismatches 50; Indels 40; Gaps 7;

QY 10 LQYAEIPKSEIPK-----VIEKAYIPVETLIKKEIPFGLNITGYTLKFLPKDIIDL 61
DB 130 MTYTVFKESGFPKRVIGSQSLDTR---FRTFVAEE-----LNIS-----VKDVTGF 175

QY 62 VKGGIADSLIEIGTSYTHAI--LPPLPSRVEAQVORDREVKEELFELSPKG--FWLPE 117
DB 176 VLGGHGDMDVPLVRSYAGGIPLEKLPKDRDLDAIVETRKGGGEIVNLLGNSAYTAPA 235

QY 118 LAYDPIIPAILKDN-----GVEYLF 137
DB 236 ASLVEMVEAILKDQRRILPAIAYLEGEVGYEGY 269

RESULT 6
US-08-838-418-2
; Sequence 2, Application US/08838418
; Patent No. 5744342
; GENERAL INFORMATION:
; APPLICANT: Sogabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,418
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 164701/1993
; FILING DATE: 02-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Allen E.
; REGISTRATION NUMBER: 37354
; REFERENCE/DOCKET NUMBER: 78339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)3533
```

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-838-418-2

Query Match      5.58; Score 104; DB 1; Length 329;
Best Local Similarity 26.08; Pred. No. 0.011;
Matches 40; Conservative 24; Mismatches 50; Indels 40; Gaps 7;

QY 10 LQYAEIPKSEIPK-----VIEKAYIPVETLIKKEIPFGLNITGYTLKFLPKDIIDL 61
DB 130 MTYTVFKESGFPKRVIGSQSLDTR---FRTFVAEE-----LNIS-----VKDVTGF 175

QY 62 VKGGIADSLIEIGTSYTHAI--LPPLPSRVEAQVORDREVKEELFELSPKG--FWLPE 117
DB 176 VLGGHGDMDVPLVRSYAGGIPLEKLPKDRDLDAIVETRKGGGEIVNLLGNSAYTAPA 235

QY 118 LAYDPIIPAILKDN-----GVEYLF 137
DB 236 ASLVEMVEAILKDQRRILPAIAYLEGEVGYEGY 269

RESULT 7
US-08-418-893D-23
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROSSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

US-08-418-893D-23

Query Match 4.7%; Score 88.5; DB 1; Length 2089;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLOVAIEPKSEIPKVEKAVIPVETLIK--EEIP-----FGLNITGVTLPKPKDIIDL 61
DB 1057 SFQADVPKAD--RVTRQGFSSVIDDASKFAQOLPELNSFGSKIAGDASKEGPNVNLQV 1114
QY 62 VKGSIASDL-IEII--GTSYTHAILPLLSRVEAQVQDRVKEELPELSPKGFWLPEL 118
DB 1115 --GALSGDISIELEKATSANKDKLNLGVRTVTALIPRKK-----DPSYYSFFQC 1164
QY 119 A---YDPIIPAILKDNQYELFADGEAMLFSAHLSAIPKIPLYPHLIKAOR-EKRF-- 172
DB 1165 SGFKEDPL-----RGMRTFHHLELGRLEENFAL 1195
QY 173 -----RYISYLLGLREL--RKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGR 223
DB 1196 ERIPAVGRNVQIVVSGSEKTARRNAQVVF-----LRAISHTPGLTTFSGARRALLQGLDE 1250
QY 224 LPLMNPKKVASWIEDKDNILLYG-----TDI-----EFIGYRDIAGYRMSVEGLEVID 272
DB 1251 LERAQANSKVS-VQSSSRIYLSLPEQSDATPEEIAKEFEG-----VID 1293
QY 273 ELNSELCLPSLKHSGRELYLRTSSWAPDKSLRIWREDE-GNAR---LNMLSYNMRGE 326
DB 1294 KLSKRL-----AQLTKLRVDE--IETKVRVTQDEDEGSPRVVPLVASSMQGE 1341

RESULT 8

US-08-418-893D-24
; Sequence 24, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLFEGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:

INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-418-893D-24

Query Match 4.7%; Score 88.5; DB 1; Length 2089;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLOVAIEPKSEIPKVEKAVIPVETLIK--EEIP-----FGLNITGVTLPKPKDIIDL 61
DB 1057 SFQADVPKAD--RVTRQGFSSVIDDASKFAQOLPELNSFGSKIAGDASKEGPNVNLQV 1114
QY 62 VKGSIASDL-IEII--GTSYTHAILPLLSRVEAQVQDRVKEELPELSPKGFWLPEL 118
DB 1115 --GALSGDISIELEKATSANKDKLNLGVRTVTALIPRKK-----DPSYYSFFQC 1164
QY 119 A---YDPIIPAILKDNQYELFADGEAMLFSAHLSAIPKIPLYPHLIKAOR-EKRF-- 172
DB 1165 SGFKEDPL-----RGMRTFHHLELGRLEENFAL 1195
QY 173 -----RYISYLLGLREL--RKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGR 223
DB 1196 ERIPAVGRNVQIVVSGSEKTARRNAQVVF-----LRAISHTPGLTTFSGARRALLQGLDE 1250
QY 224 LPLMNPKKVASWIEDKDNILLYG-----TDI-----EFIGYRDIAGYRMSVEGLEVID 272
DB 1251 LERAQANSKVS-VQSSSRIYLSLPEQSDATPEEIAKEFEG-----VID 1293
QY 273 ELNSELCLPSLKHSGRELYLRTSSWAPDKSLRIWREDE-GNAR---LNMLSYNMRGE 326
DB 1294 KLSKRL-----AQLTKLRVDE--IETKVRVTQDEDEGSPRVVPLVASSMQGE 1341

RESULT 9

US-09-141-135-2
; Sequence 2, Application US/09141135
; Patent No. 5981729
; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; APPLICANT: LEE, Jong Hun
; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisi
; TITLE OF INVENTION: Acid Isolated from Arabidopsis thaliana
; FILE REFERENCE: 1942/31
; CURRENT APPLICATION NUMBER: US/09/141,135
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 2
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-141-135-2

Query Match 4.6%; Score 87; DB 2; Length 235;
Best Local Similarity 21.0%; Pred. No. 0.39;
Matches 49; Conservative 37; Mismatches 71; Indels 76; Gaps 11;

QY 165 KAQREKFRYISYLLGLRELKKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRL 224
DB 27 KSNQKRFN-----EEQIKSLELIFESETLEPRKKVQ-----VARELG----- 65
QY 225 PLMNPKKVASWIEDKDNILLYG-----NIIYGTDTIEFIGYRDIAG---YRMSVEGLL 268
DB 66 --LQPRQMTIWFQNKRAWRKTKQLEKEVNTLR-----ANYNNLASQFEIMKKQSLV 116
QY 269 EVIDELNSELCLPSLKH-----SGRELYLRTSSWAPD-KSLRIWREDEGNARLNMLSYN- 322

RESULT 11
US-08-248-021A-2
; Sequence 2, Application US/08248021A
; Patent No. 5648240
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Jonsson, Klas
; APPLICANT: Patti, Joseph M.
; APPLICANT: Gurusiddappa, Sivashankarappa
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America

```

RESULT 12
US-09-235-451-36
? Sequence 36, Application US/09235451
? GENERAL INFORMATION:
? APPLICANT: Julius, David J.
? APPLICANT: Caterina, Michael J.
? APPLICANT: Brake, Anthony J.
? TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
? TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
? TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
? FILE REFERENCE: 9076/084CIP
? CURRENT APPLICATION NUMBER: US/09/235,451
? CURRENT FILING DATE: 1999-01-22
? PRIOR APPLICATION NUMBER: 60/072,151
? PRIOR FILING DATE: 1998-01-22
? PRIOR APPLICATION NUMBER: 08/915,461
? PRIOR FILING DATE: 1997-08-20
? NUMBER OF SEQ ID NOS: 48
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 36
? LENGTH: 764
? TYPE: PRT

```

[illegible]

; ORGANISM: A

b. 306 ADI

[illegible]

```
Db 353 PDRPEERSKVISALNLTWIEDP-SLSFSINSYSDELEISLYGLTQKEIITQLLEERFSVK 411
QY 346 -RLDAFRAIYND 356
Db 412 VHFDEIKTIYKE 423

RESULT 15
US-08-161-999-3
: Sequence 3, Application US/08161999
: Patent No. 5674733
: GENERAL INFORMATION:
: APPLICANT: Salyers, Abigail A.,
: APPLICANT: Shoemaker, Nadja B.,
: APPLICANT: Nikolich, Mikeljon P.
: TITLE OF INVENTION: Method and Materials For
: TITLE OF INVENTION: Introducing DNA Into Prevotella ruminicola
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: William Brinks Olds Hofer Gilson and Lione
: STREET: P.O. Box 10395
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60610
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM XT
: OPERATING SYSTEM: MS-DOS 3.31
: SOFTWARE: WordPerfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/161,999
: FILING DATE: 02-DEC-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/718,535
: FILING DATE: 05-JUN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Crook, Wannell M.
: REGISTRATION NUMBER: 31,071
: REFERENCE/DOCKET NUMBER: 3617/22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 321-4200
: TELEFAX: (312) 321-4299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 641 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: US-08-161-999-3
```

```
Query Match 4.4%; Score 83.5; DB 1; Length 641;
Best Local Similarity 19.6%; Pred. No. 4.6;
Matches 73; Conservative 53; Mismatches 143; Indels 103; Gaps 16;

QY 24 IEKAYIPVIEFLKEIPFGLNITGYTLKPLKDIIDLKVGGIASDLI----EIIGTSTY 79
Db 116 LQKLQIPTIIFINK-----IDRAGVNLERLYLDI----KANLSQDVLEFMQNVVDCSVYP 165
QY 80 HAILPLPLSRVEAQVORDREVKEELF---ELSPKGFWLPPELAYDPIIAILKDNCGEYL 136
Db 166 VCSQTYIKKEEYKFCVCHNDNILERLYLADSEISPADYW-----NTIIALVAKAKYIPVL 219
QY 137 FADGEAMLFSAHLNSAIKPI-----KPLYPHLIKAQREKFRYSYVLLGLREL 185
Db 220 --HGSAM-FNIGINELLDATISFTLPPASVSNRLSSVLYKIEHDPKHKRSTL----- 269
QY 186 KAIKLVFEGKVTLVKAVKDIEAPVAVNTAVMLGIGRPLMNPKKVASWIEDKNILLY 245
Db 270 ----KIIDGSLRLRDVVRINDSEKFIKIKLKTINOGR----- 303
```

```
QY 246 GTDIEFIFGYRDIAGYRMSVEGLLEVIDE-----INSELCLPSELKHSRGRELYLRTSSWA 299
Db 304 EINNVDEVGANDIA-----IVEDMDDFRIGNYLGAEPCLIQGLSHOHPAL---KSSVR 352
QY 300 PKXSLR-----IWREDESGNARLNLMSYMMGELALLAENSARGWEPLPER----- 345
Db 353 PDRPEERSKVISALNLTWIEDP-SLSFSINSYSDELEISLYGLTQKEIITQLLEERFSVK 411
QY 346 -RLDAFRAIYND 356
Db 412 VHFDEIKTIYKE 423
```

Search completed: June 2, 2002, 18:06:15
Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:03:43 ; Search time 20.94 Seconds
(without alignments)

Title: US-09-886-400-4

Perfect score:

Sequence: 1 LRALVFHGNLQYAEIPKSEI.....RRLLDAFRAIYNDWRGENGEP 364

scoring table: BLOSUM62

scoring code: `EESScore2`
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Minimum	DB seq	length: 0
Maximum	DB seq	length: 2000000000

```

MAXT||||| DB SEQ LENG CH: 2000000000

```

Post-processing: Minimum Match 0%

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing 1

Database : DTP 71.*

```
Database :
PIR_1:
1: pir1:
```

[illegible]

2: bid
3: bid

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	1494	79.6	364	2	E71144
2	169	9.0	529	2	hypothetical prote
3	162.5	8.7	529	2	hypothetical prote
4	161.5	8.6	633	2	hypothetical prote
5	157.5	8.4	686	1	probable alpha-amylase A (EC 3.2.1.1)
6	147	7.8	527	2	ALDYAT
7	145.5	7.8	655	2	uncharacterized co
8	141.5	7.5	467	2	alpha-amylase (or
9	141.5	7.5	649	2	alpha-amylase (EC
10	130.5	7.0	560	2	alpha-amylase (EC
11	138	6.8	923	2	hypothetical prote
12	124	6.6	1362	2	hypothetical prote
13	123.5	6.6	742	2	amylomullulanase P
14	120.5	6.4	324	1	hypothetical prote
15	119	6.3	602	2	methanol dehydroge
16	113	6.0	1069	2	hypothetical prote
17	112	6.0	312	2	hypothetical prote
18	111	5.9	526	2	malate dehydrogena
19	109.5	5.8	447	2	conserved hypotHet
20	106	5.6	526	2	alpha amylase [imp
21	105.5	5.6	902	2	hypothetical prote
22	104	5.5	704	2	conserved hypotHet
23	103.5	5.5	314	2	hypothetical prote
24	101.5	5.4	705	2	malate dehydrogena
25	100.5	5.4	266	2	vacu protein (ribo
26	100.5	5.4	619	2	probable xylanase/
27	99	5.3	4845	2	probable alpha-amylase repeat contain
28	98.5	5.2	588	2	BIR repeat contain
29	98	5.2	503	2	two-component resp
30	98	5.2	503	2	hypothetical prote

ALIGNMENTS

RESULT 1

E71144
hypothetical protein PH0368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000
C:Accession: E71144
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; M01D:98344137
A:Accession: E71144
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-364 <RAW>
A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BA029442.1; PID:dl030385; PID:g32
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0368
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Query Match 79.6%; Score 1494; DB 2; Length 364;
Best Local Similarity 75.8%; Pred. No. 2.7e-105;
Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

QY 1 LRALVFGNLOYAEIPKSETPKVIKAYIPVETLIKKEIPFGLNITGYLTKLPKDIID 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MRALVFGNLOYAEIPKHEISKVIEKSYFTISLIKREIPFGLNITGYSLQFPQELIH 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LVKGGIASDLTEIGTSYTHAIIPLPLSRVEAQVDRREVKEELFELSPKGWLPPLAY 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 LKREGIESELTEILGTSYTHAIIPLPLLSLREIAQIKRDRRIKEIFEVSPGGFWLPPLAY 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 DPTIPAILKONGYEYLFADGEAMLFSAHLSAISKPIKPLYPHLTKAQRERFRYISVLLG 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 DPTIPAILRDNEYEYLFADGEAMLFSAHLSAISKPIKPLYPHLTKAQRGGFYVNLILG 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 LRELKAIKLVFEGKVTLKAVKDIEAPVPMVAVNTAYMLGIRLPLMNPKNKVASWIEDKD 240
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 LRELKAINLTFGSKVTLEAVKDIEAPVPMVAVNTAYMLGIRLPLMNPKNKVASWIEDKD 240
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 NILLYGTDTIEFGYRDYAGVMSVEGLEIVDELNSLCLPSELKHSGRELYLRTSSWAP 300
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 EILLYGTDTIEFLGYRSIAGHKIAISGLMEVNLGGLCLPDRIRHNGRRLYLRTSSWAP 300
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 DKSURIWREDEGNARLNMLSYNMGELALLAENSARDARGEWPELPRRLDAPFATYNDWRGE 360
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 301 DKSURIWREDEGNARLNMLTYCMDCGEAFLAENSARDARGEWPELPRRLDAPFATYKYWNE 360
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 361 NCE 363

```
C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76831
C:Species: Pyrococcus horikoshii
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76831
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <KAN>
A:Cross-references: EMBL:D0916; GB:AB001339; NID:gi653715; PIDN:BAA18743.1; PID:gi165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match      8.7%; Score 162.5; DB 2; Length 529;
Best Local Similarity 21.4%; Pred. NO. 0.00017;
Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

QY 3 ALVPHGNLQVAEIPKSE-----IPKVIKAYIPV---ETLIKEIPFGLNIT----- 47
DB 7 ALVLHAHLPEVRHPESDYVLEEWLYEATITETIPLIHVFEGKRGDGVDFKMTSMTPL 66
QY 48 -----GYTLKFLPK----- 56
DB 67 VSMRLDPLQRYEAHLSELQELLAKEIVRNEHNGHLQYLAIFYAKFAAIRETWERYDG 126
QY 57 DIIDLKGGIADSLIEIGTSYTHAILPLPL--SRVEAQVQDRVKKEELFELSPKGF 114
DB 127 DLVTAFOFQDSNNLEITTCGATHGYLPLMKMYPQAVWAGIQVACEHYEENFGSPKGIW 186
QY 115 LPELAYDPIIPAILKONGYEYLFADGAMLFSA-----HLSNS 151
DB 187 LPECAYEGYGLERMLADAGLRYFLVDGCHGILYARPRPKFGTYPIFTGTGVAAFGRDHES 246
QY 152 -----AIPKIPPLPHLIK-AQREKFRYISYLLGLRELKAKLVF-----EGKVT 197
DB 247 QQVWSQVGPVPGVREFYKDLGWEAEYIKPYIMPNGQRNKIGIKYHKITSRDGGLS 306
QY 198 LKAVKDIEAVPVVAVNTAV-----MLGI-GRPLMNPKKVAS-----WI 236
DB 307 EKAWYD---PYNAKAKAEHASNEMYNRQQVGHLSGIMGRPPLVSPYDAELFGHMY 362
QY 237 E-----DKDNI-----LLYGTDFEPI-----GYRDIAGY----- 260
DB 363 EGPWFIDYLFKRSWFDQDTFTMTLADYLRGNPHQVCRPSOSSWGYKGFHEYWLNDTNA 422
QY 261 -----RMSVEGLLEVIDELSELCLPSELKHSGRELYL-RTSSWA 299
DB 423 WIYPHLHKAERMIELSHREAVDELEK-----ALNQAARELLLAQSSDWA 468

RESULT 4
G71241
probable alpha-amylase - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76831
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <KAN>
A:Cross-references: EMBL:D0916; GB:AB001339; NID:gi653715; PIDN:BAA18743.1; PID:gi165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match      9.0%; Score 169; DB 2; Length 529;
Best Local Similarity 21.0%; Pred. NO. 5.6e-05;
Matches 104; Conservative 55; Mismatches 162; Indels 174; Gaps 21;

QY 3 ALVPHGNLQVAEIPKSE-----IPKVIKAYIPV---ETLIKEIPFGLNIT----- 44
DB 7 ALVLHAHLPEVRHPESDYVLEEWLYEATITETIPLIHVFEGKRGDGVDFKMTSMTPL 66
QY 45 -----NITGYLKFLPK----- 56
DB 67 VSMRLDPLQRYEAHLSELQELLAKEIVRNEHNGHLQYLAIFYAKFAAIRETWERYDG 126
QY 57 DIIDLKGGIADSLIEIGTSYTHAILPLPL--SRVEAQVQDRVKKEELFELSPKGF 114
DB 127 DLVTAFOFQDSNNLEITTCGATHGYLPLMKMYPQAVWAGIQVACEHYEENFGSPKGIW 186
QY 115 LPELAYDPIIPAILKONGYEYLFADGAMLFSA-----HLSNS 159
DB 187 LPECAYEGYGLERMLADAGLRYFLVDGCHGILYARPRPKFGTYPIFTGTGVAAFGRDHES 245
QY 160 -----YPLIKAKR-----EKFRYISYLLGLRELKAKLVFEGKVT----- 197
DB 246 SQVWSSEVGPVPGVREFYKDLGWEAEYIKPYIMPNGQRNKIGIKYHKITSRDGGLS 304
QY 198 --LKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDKNIL--LYGT----- 247
DB 305 LSDKALYD---PYWAKEKAA-----EHAANFMYNRQAEHLYGIMORPPI 347
QY 248 -----DIEFIGYRDIAG-----YRMSVEGLLEVIDELNSP-----LCLPSELKH 286
DB 348 IVSPYDAELFGHWWEGVFWFDYLFKRSWDQGYTHADYLRNEPTQVCRPSQ--- 404
QY 287 SGRELYLRTSSWAPDKSLRIWREDEGNA-----RLNMLSYNNMRELALLAENSARGWEPL 342
DB 405 -----SSWGYKGFHEYWL-NETNAWIYPLHLKAAERMI-EISTL-EPDELGWRL 452
QY 343 PERRLDAPRAIYNDW 357
DB 453 NQAARELLLAQSSDW 467

RESULT 3
S76831
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
```

A:Gene: PH0193
C:Superfamily: Dictyoglomus thermophilum amylase A

Query Match 8.6%; Score 161.5; DB 2; Length 633;
Best Local Similarity 23.4%; Pred. No. 0.00027;
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;

QY 8 GNLYQAEIPKSEIPKVIKAYIPVETLIKKEIP---PGLNITGYTLKELPK---DIIDL 61
DB 17 GNFEWI-----IKRAYEKAYRPFLETL---EYPNKMVAVHISGVLVWELRNRPYIDL 68
QY 62 VKGIASDLIEIGTSYTHAILPLLSRVEAQVQDRVKEKELFELS-----PK 111
DB 69 LKSLIKQVELVAGVEYELVAIP-----EDRVEQIKLSKGWARKWGEAR 117
QY 112 GFWLPELAYOPIIPAILKNDNGEYELFADGEAMLFSAHLSA-IKPIKLYPHILKRAQREK 170
DB 118 GLMTERWNEPVLKTLREAGIEYILD-----DYHFSAGLSKEELFWPYTENGSEA 171
QY 171 RFRY-----ISYLLGLRELKAKLVF-----EGKVTLKAVKDIEAVPVWAVNTAVML 219
DB 172 IVVPIDEKLYLPPFPVNETLEYLHSLADEDESKVAV-FHDDGKFGAWPTHELVI- 229
QY 220 GIGRLPLNPKKVASWI-----EDKDNILLYGTDFIEFG-----YNDIAGY-R 261
DB 230 -----ERGWLKEPFDRISSDKINMLYS---EYLSKFRPKGLVLYPLIASYPE 274
QY 262 MSVBGL-----LEVIDELNSLCLPSKLSGRELXYLRTSSWAPDKSLRIWREDEGN 313
DB 275 MSEWSLPAQAKLFFEFIKKL-KELNLFERY-----IFVRGIG---KNF-LYKYPEGN 324
QY 314 ---ARLNLMSYMRGELALLAENSARGWEPLPERRLDAPRAYND--WRGENG 362
DB 325 YMHKRMMLLS-----KLLRNN-----PTARIFVLRAQCNDAYWGVFG 362

RESULT 5

ALDYAT
amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum
C:Species: Dictyoglomus thermophilum
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
C:Accession: S00628; A34969
R:Fukushima, S.; Kamizono, A.; Horinouchi, S.; Beppu, T.
Eur. J. Biochem. 174, 13-21, 1988
A:Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic
A:Reference number: S00628; MUID:88225097
A:Accession: S00628
A:Molecule type: DNA
A:Residues: 1-686 <FUK>
A:Cross-references: EMBL:X07896; NID:g2688; PIDN:CAA30735.1; PID:g2689
A:Accession: A34969
A:Molecule type: protein
A:Residues: 2-13 <FUK2>
C:Genetics:
A:Gene: amya
C:Superfamily: Dictyoglomus thermophilum amylase A
C:Keywords: glycosidase; hydrolase
F:2-686/Product: amylase A #status experimental <MAT>

Query Match 8.4%; Score 157.5; DB 1; Length 686;
Best Local Similarity 22.5%; Pred. No. 0.0006;
Matches 71; Conservative 53; Mismatches 110; Indels 81; Gaps 15;

QY 8 GNLYQAEIPKSEIPKVIKAYIPVETLIKKEIP-EIIPFGLNITGYTLKELPK---DIIDLVK 63
DB 18 GNFDV-----IERAVEMSKPLINEFFHPDPPINVHFSGLLWLEKNHPEYEKELK 71
QY 64 GGIASDLIEIGTSYTHAILPLLSRVEAQVQDRVKEKELFELSPKGFPLPELAYDI 123
DB 72 IMARGQIEFVSGGFYFPIIDPKQVQIKLNKIYDKFGQTPKGMWLAERWNEPH 131

QY 124 IPAILKNDNGEYELFADGEAMLFSAHLSAIIKPIKLYPHILKRAQREKRPYISYLLGLRE 183
DB 132 LVKYLAEAGIEYVVD--DAHFFSVGL-----KEEDLFGY--YLMESQG 171
QY 184 LRKAIKLVFEKGVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWI-----ED 238
DB 172 YKLA--VFPISMKLRYL-----IPFADPEETITVLDKFASED 206
QY 239 KDNILLYGTDFIEFG-----YRDIAGYRMSVEGLLEV-IDELNSELCLPS-----E 283
DB 207 KSKTALLFDDGKFGMLWPDYRTV--YE---EGWLETFVSKIKENFLLVTPVNLTYMQR 261
QY 284 LKHSGRELYLRTSSW 298
DB 262 VKPKGR-IYLPASY 275

RESULT 6
F97197
uncharacterized conserved protein, related to alpha-amylase/alpha-mannosidase CAC2414
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: F97197
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80369.1; PID:g15025430; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2414
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 7.8%; Score 147; DB 2; Length 527;
Best Local Similarity 18.9%; Pred. No. 0.0026;
Matches 99; Conservative 62; Mismatches 120; Indels 212; Gaps 20;

QY 5 VFHGNLYQAEIPKSEIPKVIKAYIPVETLIKKEIPFG-----DL 61
DB 9 VLHSHMPFVRHPETROSLERWLFEMSECYIPLIEYDNLKDNKIFRTWMSITPPLMS 68
QY 44 -----LNITGYTLKELPKDII-----DL 61
DB 69 MLODEYLSRYLNVLYKKTIELSEKEILRTKNREKNKVALFYNKRAENTLKIEKYDNL 128
QY 62 VKGIASD---LIBIOTSYTHAILPLLSR--VBAQVORDREVKEELFELSPKGFWLP 116
DB 129 INAFKRYDLRGVEIITCAATHALLPLILNQAVKQATIGVQSYINTMGHEPNGLWP 188
QY 117 ELAYDPIIPAILKNDNGEYELFADGEAMLFSAHLSAIIKPIKLYPHILKRAQREKRPYIS 176
DB 189 ECATYIGIDNLSEFGIKYFISGKAIDYASP-----KPY----- 224
QY 177 YLLGLRELKRAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRLPLMNPKKVASWI 236
DB 225 -----GTNT-----PIAAPSGVCAFG 240
QY 237 EDKNILLYGTDFIEFG-----YRDIAGYRMSVEGLLEVIDELNSELCLPSLKH 286
DB 241 RDMDS--SYQWSDFMGPGDFNVREYRDI-GFELPMEYIKPYINE--NGRIDTGFYI 295
QY 287 -----SGRE-LYLRTSS---WAPDKSLRIWREDEGN-ARLNN----- 318
DB 296 YKITSNGEKGIYRKNAMKKVWEHSHASCHRDQINAAANAKDKPPIICPYDTELYG 355
QY 319 -----LSYMRGELALLAENSARGWEPLPERRLDAPRAYNDW 357

Db 356 HWFEGPDPFINAFIRKSAEDWTSELYITPTLYLKNNSWVQCSPSP-----SSW 404

QY 358 RGENGE 363

Db 405 -GENGD 409

RESULT 7

E75206

A:Accession: B64501

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-467 <BUL>

A:Cross-references: GB:067601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212

C:Genetics:

A:Map position: FOR1586396-1587799

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A:Pathway: glycogen/starch degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 141.5; DB 2; Length 467;

Best Local Similarity 22.8%; Pred. No. 0.0056;

Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;

QY 14 EIPKSEIPKVIKAVIPVETLIK--BEIPFGLN--ITG----YTLKFLPKD-IIDLKVG 64

Db 34 KLNKEVFNKVKCYIPTNEILLLEIDEDYFKVNSITGVFEQALEF--NDYVLDLFD 91

QY 65 GIASDLIEIIGTSTHAILPL--SRVEAOVDRDREKEELFELSPKGFELAYDPI 123

Db 92 LVKTGNVELIAETVHSLTSLFETEDFIEDIEMHRKMYKEIFGAKVFRNTELYNNR 151

QY 124 IPAILKONGYEVLPADG--EAMLFSAHLNSAIKPIKPLYPHLIKAQRKRPYIS---YLL 179

Db 152 IAKIADLGFKAIFTEGIEKIL-----GWSRPNLYQSPDGMKI 190

QY 180 GLRELRKAIKLVFECKVTLKAVKDIEAVP-----VWAVNTAVMLGIGRLPLMPPKV 232

Db 191 LLRNRLSDDIGR-----FSARDWDQYPLTADYAIWLASTFGEVINI----YMDYETF 241

QY 233 AS--WIEDKDNILLYGTDI-EFIGYR--DIAGY-RMSVEGLLEVIDELNSELCLPSELKH 286

Db 242 GEHNKE-----IGIFEFLYLPETIAKHEHELVNVVSEVVDLEPR----- 283

QY 287 SCORELYLR---TSSWA-PDKSLRTWRDEGNARLNMGLSYNMRGELA--LLAENSADARGWEP 341

Db 284 --GEIYVHEFATISWADTERDVSAWL---GN-KMORISFEKLKIDIGKIFENSKL---- 333

QY 342 LPERLDFAFRAYINDWR 358

Db 334 ---KKLNKFEDELYKMYK 347

RESULT 9

A49512

alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus

C:Species: Pyrococcus furiosus

C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000

C:Accession: A49512

R:Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, J. Biol. Chem. 268, 24402-24407, 1993

A:Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus

A:Reference number: A49512; MUID:94043280

A:Accession: A49512

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-649 <LAD>

A:Cross-references: GB:L22346; NID:g347939; PIDN:AAA72035.1; PID:g347940

C:Genetics:

A:Gene: amya

A:Start codon: GTG

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A:Pathway: glycogen/starch degradation

C:Superfamily: Dictyoglomus thermophilum amylase A

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 141.5; DB 2; Length 649;

Db 356 HWFEGPDPFINAFIRKSAEDWTSELYITPTLYLKNNSWVQCSPSP-----SSW 404

QY 358 RGENGE 363

Db 405 -GENGD 409

RESULT 7

E75206

A:Accession: B64501

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-467 <BUL>

A:Cross-references: GB:067601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212

C:Genetics:

A:Map position: FOR1586396-1587799

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A:Pathway: glycogen/starch degradation

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 141.5; DB 2; Length 467;

Best Local Similarity 22.8%; Pred. No. 0.0056;

Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;

QY 14 EIPKSEIPKVIKAVIPVETLIK--BEIPFGLN--ITG----YTLKFLPKD-IIDLKVG 64

Db 34 KLNKEVFNKVKCYIPTNEILLLEIDEDYFKVNSITGVFEQALEF--NDYVLDLFD 91

QY 65 GIASDLIEIIGTSTHAILPL--SRVEAOVDRDREKEELFELSPKGFELAYDPI 123

Db 92 LVKTGNVELIAETVHSLTSLFETEDFIEDIEMHRKMYKEIFGAKVFRNTELYNNR 151

QY 124 IPAILKONGYEVLPADG--EAMLFSAHLNSAIKPIKPLYPHLIKAQRKRPYIS---YLL 179

Db 152 IAKIADLGFKAIFTEGIEKIL-----GWSRPNLYQSPDGMKI 190

QY 180 GLRELRKAIKLVFECKVTLKAVKDIEAVP-----VWAVNTAVMLGIGRLPLMPPKV 232

Db 191 LLRNRLSDDIGR-----FSARDWDQYPLTADYAIWLASTFGEVINI----YMDYETF 241

QY 233 AS--WIEDKDNILLYGTDI-EFIGYR--DIAGY-RMSVEGLLEVIDELNSELCLPSELKH 286

Db 242 GEHNKE-----IGIFEFLYLPETIAKHEHELVNVVSEVVDLEPR----- 283

QY 287 SCORELYLR---TSSWA-PDKSLRTWRDEGNARLNMGLSYNMRGELA--LLAENSADARGWEP 341

Db 284 --GEIYVHEFATISWADTERDVSAWL---GN-KMORISFEKLKIDIGKIFENSKL---- 333

QY 342 LPERLDFAFRAYINDWR 358

Db 334 ---KKLNKFEDELYKMYK 347

RESULT 9

A49512

alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus

C:Species: Pyrococcus furiosus

C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000

C:Accession: A49512

R:Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, J. Biol. Chem. 268, 24402-24407, 1993

A:Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus

A:Reference number: A49512; MUID:94043280

A:Accession: A49512

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-649 <LAD>

A:Cross-references: GB:L22346; NID:g347939; PIDN:AAA72035.1; PID:g347940

C:Genetics:

A:Gene: amya

A:Start codon: GTG

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

A:Pathway: glycogen/starch degradation

C:Superfamily: Dictyoglomus thermophilum amylase A

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 7.5%; Score 141.5; DB 2; Length 649;

Db 235 RYGNILPA-KTKRSTLRPVF-----LKNGIAVFARNRET-----GIQWWSA 274

RESULT 11

G83826

hypothetical protein BHL415 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: G83826

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; PMID:20512582; PMID:11058132

A:Accession: G83826

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-923 <STO>

A:Cross-references: GB:AP001512; GB:BA000004; NID:gi10174030; PIDN:BA05134.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BHL415

Query Match 6.8%; Score 128; DB 2; Length 923;

Best Local Similarity 20.1%; Pred. No. 0.16;

Matches 97; Conservative 60; Mismatches 167; Indels 158; Gaps 24;

QY 3 ALVPHGNLQVAEIPKSE-----IPKVEKAVIPYETLIKEEIPFGLNI----- 46

Db 7 SLVLHAHLPVYRHOEDRLEERLWFLEAMSETVYPLLWALEKLPVKHAVTISTPTPPVMEML 66

QY 47 -----TGY-----TLKFLPKD-----IIDLVKG-----G 65

Db 67 SDPLVQRYLNLHLENTQLLKKEKRTNDQTNLYQYKQYKELKATFLQWDRNLLIG 126

QY 66 IASDLIE-----IIGSYTHAILPLPLSLR-VEAQVQDREYKEBELFELSPKGFVLPDLA 119

Db 127 FRS-LMENEQTLMTSAATHAFPPYIKTEKRAQVHGIACFEQHFQKKGKPLGFWLPSCA 185

QY 120 YDPIPAILKNDNGEYLFADGEAMLF---SAHLNSAIKPKLYLPH---LIKAQREKRF 172

Db 186 FSPQVDRIEFGIRYTFVDEHAVALTADTPHKGSS---APIYSPHGIALPRTTSLA 241

QY 173 RYTSYLLGLRELKAKLVFEGKVTLKA-VKDIEAVPVW-----VAVNTAVMLG 220

Db 242 KWSSTILG-----YPGVDYREFYRDYADREWDYKPHVHKDGIIDTG--LK 288

QY 221 IGRPLMNPKK---VASWIEDK-----DNILLYGTDIEFI 252

Db 289 YHRITGHTTEKDLVYREWAERKQVQEHANHFHIGAIHHEIDQHGQNFPPYVMVTPFDALF 348

QY 253 GYRDIAGYRMSVGLLEV-----IDELNSCLPSELKHSRGLYLTSSWAPDKSLRW 308

Db 349 GHMWFEGPEW-IEALYEQGADRVSFITPELYLQRYQ-QFQTAHVSFTWGRDGVGHVWL 406

QY 309 EDEGNARLNMLSYNMRGELALL-----AENS DARGWEPLPERLLDAFRAYN 355

Db 407 NDH-NAMWYRHYHMEKDLAKIVAMYPPQVTVLEKQAIQQWREW-----MLAVSS 455

QY 356 DW 357

Db 456 DW 457

RESULT 12

A75207

amylopullulanase PAS0122 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999

C:Accession: A75207

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

Best Local Similarity 22.8%; Pred. No. 0.009;

Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

QY 25 EKAVIPVETLKEEIP---FGLNITGYTLKFLP---KDIIDLKVGGIASDLIEIGTSY 78

Db 29 EKCWPFLETL---EYVPMKVAIHTSGPLIEWLQDNRPYIDLLSLVKRGQVEIVVAGF 86

QY 79 THAILPLPLSRVQAOVQDREYKE---ELFEISPKGFVPELAYDPIIPAILKNDNGEY 136

Db 87 YEPVLASIP---REDRTIQLKWEKASIGSFGDARGVWLTQVWQPELVKTLKESGIDYV 143

QY 137 FADGEAMLFSAHLNSAIKPKLY-PLHLKAQRE-----KFRYSYLLGLRELKAKA 187

Db 144 IVD-----DYHMSAGLSKEELYIYFEDGGEVAVTPIDEKLR---YLIPFRPVQKV 194

QY 188 IKL---VFEQKVTLKAV-KDIEAVPVWVAVNTAVMLGIGRLPLMNPKKVASWI----- 236

Db 195 LEYLSLIDGDESKVAVFHDDGKFGIWPGETVWVY-----EKGWLFREFFDR 241

QY 237 ---EDKDNILLYGTDIE-----FIGYRDIAGYRMSVEGLEVIDELNSCLCP----- 281

Db 242 ISDEKNILMLYTELEKYPKRGVLVPLIASY-----FEM-----SEWSLPKQAKRLF 289

QY 282 ---SELKHSQ---RELYLRTSSAPDKSLRIWRDEGN---ARLNMLSYNMRGELALL 330

Db 290 VEFVNLKVGKIGEKYRVFVVGGIW---KNF-FYKYPESNYMHKRMVSKLVLRN--- 341

QY 331 AENS DARGWEPLPERLLDAFRAYND---WRG 359

Db 342 -----PEARLYLLAQCNDAVWHG 360

RESULT 10

D71011

hypothetical protein PH1386 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: D71011

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A:Reference number: A71000; PMID:98344137

A:Accession: D71011

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-560 <RAW>

A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30492.1; PID:g3257809

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1386

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match 7.0%; Score 130.5; DB 2; Length 560;

Best Local Similarity 25.6%; Pred. No. 0.049;

Matches 60; Conservative 28; Mismatches 81; Indels 65; Gaps 12;

QY 14 EIPKSEIPKVIKRAYIPVETL-----IKKEIIPFGLNITGYTLKFLPKDIDL---VK 63

Db 71 EYIKREFEKYMERKLGMEEDLERFKDEKLREAINF---MIGYF-----KDVYSYWKSID 122

QY 64 GGIASDL-----IEIIGTSYTHALPLPLSR-VEAQVQDREYKEBELFELSPKGF 114

Db 123 GNILKGFRELQDEGYEVITSAATHGVLPDLGRDEAIEAQLNGIKVYKFGYGRKPRGTW 182

QY 115 LPELAYDPI-----IPAILKNDNGEYLFADGEAMLFSAHLNSAIKPKIP 158

Db 183 LPECAIRPDGLWKSFGSTGEKWKRGKIEHFLKKGIEYFVE-----SHLIDK-GPVS 234

QY 159 LYPHLIKQREKRRPYISYLLGLRELKAKLVFEGKVTLKA-VKDIEAVPVVWA 212

A:Reference number: A75001
A:Accession: A75207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1362 <RAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CAB49104.1; PID:el51499
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: apu; PAB0122

Query Match 6.6%; Score 124; DB 2; Length 1362;
Best Local Similarity 21.5%; Pred. No. 0.55;
Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEEIPFGLNITGVTLKPKDIIDLKGGTASDLIEITIGTSYTHAILPLP--- 87
DB 236 VETVLKQHM-WLNNHT-----FEEHEKINLLG---NGNVEVTVPVTHPIGPILNDEGW 266
QY 98 LSRVAQVORDREVKEELF---ELSPKGFWLPPELAYDPIIPAILKNDGNGYELFADG---E 141
DB 287 YEDFDAQVKKANELYKEYLGAGKVTGKGGAAESALNDKTLLEILAENGKWMVTDQLVLE 346
QY 142 AMLFSAHLNSAIKP-----IKPLYPHLIKA 166
DB 347 KLGVPKTIESYKVPWAQFGDKKIYLFPRNHDSRVGFRVAGMNOYDAVKNFVEELKI 406
QY 167 QREKFRYSYLLGLRELKKAIKLVFGKVTLKAV-KDIEAVPVVAVNTAV-----ML 219
DB -407 QKONYDGSIVYITLDGENPWEHYFPDGKLFLEELYROLELQKKGLIRVTVPSEYEMF 466
QY 220 G--IGRLPLMNPKKVASMTEDKNDLILYCTDIETFIYRDIAGYRMSVEGLLEIDELNSE 277
DB 467 GOKANKLTPKMKRLDFTEDNVNALLKAKVL-----GELYDMVGYTE----- 509
QY 278 LCLPSLKHSGBRELYLRTSSWAPDKSLRIWREDEGNARLNLMLSYMRGELALLAENS-D-A 336
DB 510 -----EMQWPESSWT-DGTLSTW-----IGEPQENIAWYLYLARKALFENKDNV 553
QY 337 RGWEPLPERRLDAFRAIYNDW 357
DB 554 KDNKAYEY---LFRAGSDW 571

RESULT 13
AC1970
hypothetical protein alr1310 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC1970
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <GB>
A:Cross-references: GB:BA000019; PIDN:BA073267.1; PID:g17130657; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1310

Query Match 6.6%; Score 123.5; DB 2; Length 744;
Best Local Similarity 22.6%; Pred. No. 0.25;
Matches 79; Conservative 49; Mismatches 111; Indels 111; Gaps 19;

QY 68 SDIEIIGTSYTHAILPL-----LPLSR-----VEAQVORDREVKEELF 106
DB 208 SGOLEVTTPYTHPLPLADNNSGRVAVPNMALPESRFQNSWDIPRHLRKAWELYTERF 267

QY 107 ELSPKGFWLPPELAYDP-IIPAILKNDGNGYELFAD-----GEAMLFSAHLNSAIKPIKP--L 159
DB 268 GQEPKGLWPSEQSVSPDILPYIIR-QGFQWICSDAELVGLWTLKHFHRDGAGNVQOPELL 326
QY 160 Y-PHLIKAQ-----REKFRYSYLLGLRELKKAIKLVFGKVTLKAVKDIEAVPVW 210
DB 327 YRPYRLATPAGDLAIVFRDHR---LSDLIGF-----TYGAMPKAKAAADL 368
QY 211 VAVNPAVMLGIGIRLPLMNPKKVASWIEDKKNILLYGTD---IEFIGYRDIAGY-RMSVE 265
DB 369 VCHLOAIKAKMORERPEQWLVLTALDGENCWEEYPQDGKPFLEAL-YQSUSNEPHIKLV 427
QY 266 GLEVIDENSELCPSELKHSGBRELYLRTSSWAPDKSLRIWREDEGNARLNLMLSYMRG 325
DB 428 TVSEFIEEFPATATIPAEQLHSG-----SMV-DGSFTTWIGDP----- 464
QY 326 ELALLAENSADRGWPELPERL-----DAFRAIY-----NDW 357
DB 465 -----AKN---RANDYLFTEARILMLANHPATEENNPEAWAELYAAEGSDW 506

RESULT 14
B69553
methanol dehydrogenase regulatory protein (moxr) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B69553
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:98049343
A:Accession: B69553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <KLE>
A:Cross-references: GB:AB001108; GB:AB000782; NID:g2689431; PIDN:AAB91247.1; PID:g265
C:Superfamily: methanol dehydrogenase regulatory protein

Query Match 6.4%; Score 120.5; DB 1; Length 324;
Best Local Similarity 21.7%; Pred. No. 0.13;
Matches 71; Conservative 60; Mismatches 121; Indels 75; Gaps 14;

QY 1 LRALVFHGNLQVAEIP---KSEIPKVIKAVIPVITLKEEIPFGLNITGYTLKFLPKD 57
DB 35 LAAALTNGLIFEDYFGLGKILLAKVAVR-----IGADYRVQVTFDILLPSD 82
QY 58 IIDLVKGGI---ASDLIEII-GTSYTHAIL-----PLPLSRVAQVORDREVKEEL 105
DB 83 II-----GVKIRGDRFVFKGPIFTNVLDAEINRSPKPTQAALLEAMEEKQITVEGET 137
QY 106 FELSPKGFWLPPELAYDPIIPAILKNDGNGYELFADGEAMLFSAHLNSAIKPIKPLYPHLK 165
DB 138 FLSLMPFFVL--ATONPI-----EQETYPLEAQMDRFLMRKPGYPSIE 182
QY 166 AQREKFRYSY-----LIGLRELKKAIKLVFGKVTLKAVKDIEAVPVW 211
DB 183 EEMEILRRRISWRKDDPTEDVEPVVSLFTFRRIQDAVEAVYVDKXILKYISELVRA---T 239
QY 212 AVNTAYMLGI---GRPLMNPKKVASWIEDKKNIL---LYGTDIEFIGYRDIAGYRMSVE 265
DB 240 REHELVELGSSPRGGLALLKLARAVMDGRDFVIPDDVKRVAVEALAHRVILAFAYE 299
QY 266 GLL--EVIDENSELCPSELKHSGRE 290
DB 300 GLRAEEVVEILLNSVRP---KYEAQE 323

Search completed: June 2, 2002, 18:06:42
Job time: 179 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:05:58 ; Search time 29.22 Seconds
(without alignments)
2155.036 Million cell updates/sec

Title: US-09-886-400-4
 Perfect score: 1877
 Sequence: 1 LRALVFHGNLQVAEIPKSEI.....RRLDAFRAIYNDRGENCEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters:

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

```

1: 1: sp:archaea:*
2: 2: sp:bacteria:*
3: 3: sp:fungi:*
4: 4: sp:human:*
5: 5: sp:invertibrate:*
6: 6: sp:mammal:*
7: 7: sp:mhc:*
8: 8: sp:organelle:*
9: 9: sp:phage:*
10: 10: sp:plant:*
11: 11: sp:rodent:*
12: 12: sp:virus:*
13: 13: sp:vertebrate:*
14: 14: sp:unclassified:*
15: 15: sp:rvirus:*
16: 16: sp:bacteriap:*
17: 17: sp:archeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	1532	81.6	364	1	Q9HHB5	Q9HHB5 pyrococcus
2	1494	79.6	364	1	Q58106	Q58106 pyrococcus
3	162.5	8.7	529	16	P74630	P74630 synechocyst
4	149.5	8.0	378	17	Q97BM4	Q97BM4 thermoplasm
5	147	7.8	527	16	Q97GF3	Q97GF3 clostridium
6	142.5	7.6	443	17	Q973T0	Q973T0 sulfolobus
7	136.5	7.3	895	17	Q972N0	Q972N0 sulfolobus
8	130.5	7.0	560	17	O50094	O50094 pyrococcus
9	128	6.8	923	16	Q9K004	Q9K004 bacillus ha
10	124	6.6	1362	17	Q9V294	Q9V294 pyrococcus
11	122.5	6.5	357	17	Q9HL91	Q9HL91 thermoplasm
12	120.5	6.4	324	17	Q30246	Q30246 archaeoglob
13	119	6.3	602	17	Q9VOM7	Q9VOM7 pyrococcus
14	111	5.9	526	16	O83377	O83377 treponema p
15	109.5	5.8	447	17	Q97ZD2	Q97ZD2 sulfolobus
16	106	5.6	526	16	O53278	O53278 mycobacteri

17	105.5	5.6	902	17	Q97TY0
18	104	5.5	704	10	Q9ZVN2
19	103.5	5.5	458	10	Q9MBD0
20	103.5	5.5	701	2	Q93MG7
21	103	5.5	754	13	Q13131
22	102	5.4	754	13	Q13132
23	100.5	5.4	266	16	Q97M42
24	100.5	5.4	619	16	Q83A82
25	99	5.3	4845	11	Q88738
26	98	5.2	503	5	Q17585
27	98	5.2	764	16	Q67347
28	97.5	5.2	402	2	Q68779
29	97.5	5.2	505	16	Q92DR2
30	97.5	5.2	627	10	Q9ZQF1
31	97.5	5.2	684	10	Q9MZS4
32	97.5	5.2	888	16	Q98PQ2
33	97.5	5.2	1085	16	Q25577
34	97	5.2	510	16	Q66373
35	97	5.2	531	17	Q97BF8
36	97	5.2	610	2	Q32582
37	97	5.2	929	5	Q9VR32
38	96.5	5.1	197	13	Q9VBE8
39	96	5.1	636	2	Q87139
40	96	5.1	636	2	Q34235
41	96	5.1	1150	16	Q90UY8
42	95.5	5.1	1134	17	Q90Y58
43	95.5	5.1	1326	5	Q16928
44	95	5.1	296	2	Q92597
45	95	5.1	394	17	Q9UYC0

ALIGNMENTS

RESULT	1
ID	Q9HHB5
PRELIMINARY;	PRT; 364 AA.
Q9HHB5	
Q9HHB5	
AT	01-MAR-2001 (TRENBLrel. 16, Created)
DT	01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT	01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE	ALPHA-GALACTOSIDASE.
GALA.	
GN	Pyrococcus furiosus.
OS	Pyrococcales; Thermococcaceae; Pyrococcus.
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
NGBI_TaxID:	2261;
[1]	
RN	SEQUENCE FROM N.A.
RP	[1]
STRAIN-DSM	3638;
Verhees C.H.;	
RA	Submitted (Oct-1999) to the EMBL/GenBank/DDBJ databases.
RL	EMBL; AF195244; AAG28455.1; -.
DR	InterPro: IPR004300; Glyco_hydro_57.
DR	Pfam: PF03065; Glyco_hydro_57; 1.
SQ	SEQUENCE 364 AA; 41545 MW; 3E8EQAF5BDCCF2A5 CRG64;

```

Query Match      81.6%; Score 1532; DB 1; Length 364;
Best Local Similarity 79.1%; Pred. No. 4.9e-117;
Matches 287; Conservative 38; Mismatches 38; Indels 0; Gaps 0;

QY 1 LRALPHGNLQVAETPKSEIRPKVIEKAIPIWETLKEEIPFGUNITGYTLKFLPKDID 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MRALPHGNLQVAETPKSEIRPKVIEKAYPTISELIRREIPFGUNITGYSLFLPKDLIA 60

QY 61 LVKGGTASLIEIIIGTSTYTHAILPLPLSRVQAOQDRVKKEELFSPKGGWLPDELAY 120
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 LIKEGIESGLIEIIGTSTYTHAILPLPLSRVQAIKQDRVKENILEVSEGGWLPDELAY 120

QY 121 DPTIPAILKNDNGEYILFADGAMLTFSAHLSNAIKPKIPLYPHITKAQREKKRYISYLLG 180
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 DPTIPAILPNNVYEVLFADGAMLTFSNHSATKPKIPKLYPHITKAQGGGGLVYVNYLLG 180

```

QY 181 LRELKAIKLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
 DB 181 LRELKAINLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
 QY 241 NILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYRTSSWAP 300
 DB 241 EILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYRTSSWAP 300
 QY 301 DKSLRIWREDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDWRGE 360
 DB 301 DKSLRIWREDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDWRGE 360
 QY 361 NGE 363
 DB 361 NGK 363

RESULT 2
 ID P74630 PRELIMINARY; PRT; 364 AA.
 AC P74630
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL 41.8 KDA PROTEIN PH0368.
 GN PH0368
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 CX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kavarabavasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai K., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-i., Foshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA Res. 5:55-76(1998).
 DR EMBL: AP000002; BAA29442.1; .
 DR InterPro: IPR004300; Glyco_hydro_57.
 DR Pfam: PF03065; Glyco_hydro_57; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 364 AA; 41755 MW; 784B36A4A975B4D CRC64;

Query Match 79.6%; Score 1494; DB 17; Length 364;
 Best Local Similarity 75.8%; Pred. No. 6.3e-114;
 Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

QY 1 LRALVHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 60
 DB 1 LRALVHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 60
 QY 61 LVKGGIASDLIEIGTSYTHAILPLLSRVEAQVDRREVKEELFELSPKGFELPCLAY 120
 DB 61 LVKGGIASDLIEIGTSYTHAILPLLSRVEAQVDRREVKEELFELSPKGFELPCLAY 120
 QY 121 DPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 180
 DB 121 DPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 180
 QY 181 LRELKAIKLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
 DB 181 LRELKAINLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
 QY 241 NILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYRTSSWAP 300
 DB 241 EILLYGTDIEFGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYRTSSWAP 300

QY 301 DKSLRIWREDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDWRGE 360
 DB 301 DKSLRIWREDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDWRGE 360
 QY 361 NGE 363
 DB 361 NGK 363

RESULT 3
 ID P74630 PRELIMINARY; PRT; 529 AA.
 AC P74630
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 62.1 KDA PROTEIN.
 GN SLL0735.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 CX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Simpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL: D90916; BAA18743.1; .
 DR InterPro: IPR003803; DUF200.
 DR Pfam: PF02651; DUF200; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 529 AA; 62075 MW; 4C1A45048A784E30 CRC64;

Query Match 8.7%; Score 162.5; DB 16; Length 529;
 Best Local Similarity 21.4%; Pred. No. 7.1e-05;
 Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

QY 3 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 47
 DB 3 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 47
 QY 7 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 66
 DB 7 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKPLPKDIID 66
 QY 48 -----GYTLKPLPKDIID----- 56
 DB 48 -----GYTLKPLPKDIID----- 56
 QY 67 VMLRDLPLQOYEAHLSSLQELLAKEIVRNEHNGHLQYLAADFYAKEFAAIRETWERYDG 126
 DB 67 VMLRDLPLQOYEAHLSSLQELLAKEIVRNEHNGHLQYLAADFYAKEFAAIRETWERYDG 126
 QY 57 DIIDLKGGIASDLIEIGTSYTHAILPLLSRVEAQVDRREVKEELFELSPKGFELPCLAY 114
 DB 57 DIIDLKGGIASDLIEIGTSYTHAILPLLSRVEAQVDRREVKEELFELSPKGFELPCLAY 114
 QY 127 DLVTAFAKQFQDSNNLEITCGATHGYFPLMKMPQAVWAQIKVACEHYEENFGSRSPKGTW 186
 DB 127 DLVTAFAKQFQDSNNLEITCGATHGYFPLMKMPQAVWAQIKVACEHYEENFGSRSPKGTW 186
 QY 115 LPELAYDPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 151
 DB 115 LPELAYDPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 151
 QY 187 LPELAYDPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 246
 DB 187 LPELAYDPILPAIKLDNGEYLFADGEMFLSALNSAIKPIKPLYPHLIKAQRRFRFYISYLLG 246
 QY 152 -----AIKPIKPLYPHLIKAQRRFRFYISYLLGRELKRAIKLVF-----EGKVT 197
 DB 152 -----AIKPIKPLYPHLIKAQRRFRFYISYLLGRELKRAIKLVF-----EGKVT 197
 QY 247 QQVWSSQVGYGPDVYREFYKDLGWEAEYIKPIKPLYPHLIKAQRRFRFYISYLLG 306
 DB 247 QQVWSSQVGYGPDVYREFYKDLGWEAEYIKPIKPLYPHLIKAQRRFRFYISYLLG 306
 QY 198 LKAVKDIKAVPVVAVNTAV-----MLGI-GRPLMNPKNKVAS-----WI 236
 DB 198 LKAVKDIKAVPVVAVNTAV-----MLGI-GRPLMNPKNKVAS-----WI 236
 QY 307 EKAVD-----PYWAKEAAEHASNFNMQQQVGHLSGIMGRLPLVSPYDAELFGHWY 362
 DB 307 EKAVD-----PYWAKEAAEHASNFNMQQQVGHLSGIMGRLPLVSPYDAELFGHWY 362
 QY 237 E-----DKDNI-----LLYGTDIEFT-----GYRDIAGY----- 260
 DB 237 E-----DKDNI-----LLYGTDIEFT-----GYRDIAGY----- 260
 QY 363 EGPWFIDYLFKRSFQDQDTFHTLADYLRGNPHQVCRPSQSSWGKGFHFWYLDNTNA 422
 DB 363 EGPWFIDYLFKRSFQDQDTFHTLADYLRGNPHQVCRPSQSSWGKGFHFWYLDNTNA 422
 QY 261 -----RMSVEGLLEVIDELNSCLPSELKHSGRGLYRTSSWA 299

Db 423 WYPHLHKAERMIELSHREAVDELEK-----ALNQAAPELLLAOSSDWA 468

RESULT 4
Q97BM4 PRELIMINARY; PRT; 378 AA.

ID Q97BMA AC Q97BM4;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TremBLrel. 18, Last annotation update)
DE ALPHA-AMYLASE.
GN TVG0421416.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339; *
RN [1]_TaxID=50339;
RP SEQUENCE FROM N.A.
RX STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=T0570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Chiya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunohiba T., Yanamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaea adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium".
RL Proc. Natl. Acad. Sci.: U.S.A. 97:14257-14262(2000).
DR EMBO; AP000992; BABS9573.1.; -.
KW Complete proteome.
SQ SEQUENCE 378 AA; 44957 MW; 49FCF63BE6B3A7 CRC64;

Query Match 8.0%; Score 149.5; DB 17; Length 378;
Best Local Similarity 21.1%; Pred No. 0.00051;
Matches 66; Conservative 49; Mismatches 123; Indels 75; Gaps 11;

OY 22 KVIERAYIPVIETLIKESIFGLNITG-----YTklFLPKDIIdLYKGFIASDLIEI 73
DB 42 RVAERSYPATENLMMEYGIKSFISITAVEQALMYNTK----VDAIDDYVKSGLCGM 96
OY 74 IGSYTHAILPILLSRYEAOVRDERVEKELFELSPKFWLPDELAYDPPIPAILKDNGY 133
DB 97 LSEITYHSLSATWNDFEKRVQDMERDLMKRINFNVVPKVFNTELIYDDRIAEMVKRWGF 156
OY 134 EYLFADGEAMLFSAHNSAIRKPILPLPHLIKAORFRFYIS---YLIGLRKLAKALK 190.
DB 157 TSIIITEG-----TDSIVKHDSPNV-----RYASPSGLNLyLNRYVMSDNI 196
OY 191 VFEGKYTLKAVXDIAEPVAVNTAVMLGICRPLMNPKKVASWTEDKDNILLTGTDIE 250
DB 197 SFRESNT--KWKDYF-----LTADYAKWINESEGMDVN---L 229
OY 251 FTGYRIAGYMSVGGILLEVIDEL-----NSECLFPSELKHGRELYL--RTSSWA 299
DB 230 FMDYTFGHQTQGTFIDFKPYLVVFROYGIETITISEAEKRHRVKDVLSIPTISWA 289
OY 300 PD-KSLRIWREDE 311
DB 290 DTRRDLSANLENE 302

RESULT 5
Q97GF3 PRELIMINARY; PRT; 527 AA.

ID Q97GF3 AC Q97GF3;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-DRC-2001 (TremBLrel. 19, Last annotation update)
DE UNCHARACTERIZED CONSERVED PROTEIN, RELATED TO
DE ALPHA-AMYLASE/ALPHA-MANNOSIDASE.
GN CAC2414.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus; Clostridium group; Clostridiaceae;

OC	Clostridium.
OX	NCBL_TaxID=1488;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
RD	MEDLINE=21359325; PubMed=11466286;
RA	Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RX	Gibson R.L., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA	Tatusov R.I., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA	Bennett G.N., Koonin E.V., Smith D.R.;
RT	"Genome sequence and comparative analysis of the solvent-producing
RT	bacterium Clostridium acetobutylicum.";
RL	J. Bacteriol. 183:4823-4838(2001).
DR	EMBL: AE007742; AAK80369.1 ;
DR	InterPro: IPR003803; DUF200.
DR	Pfam: PF04651; DUF200; 1.
KW	Complete proteome.
SQ	SEQUENCE 527 AA: 61485 MW: 03706EI6907929C6 CRC64;

Query Match	7.88;	Score 147;	DB 16;	Length 527;
Best Local Similarity	18.9%;	Pred. No. 0.0013;		
Matches	92;	Conservative	62;	Mismatches 120; Indels 212; Gaps

QY	5	VFHGNLOVAEIPKSE-----IPKVIEKAYIPVTE--TLIKKEIPFG-----	43
Db	9	VLHSHMPFVRHPETRDSLEERWFPEAMSCYPLEIVDNLKKONIFRMTSITPPLMS	68
		::: ::: ::: ::: ::: ::: ::: :::	
QY	44	-----LNITGYTLKFLPKDII-----	-DL 61
		::: ::: ::: ::: ::: ::: :::	:
Db	69	MLQDEVLSRVLYNLKKTIELSEKEILTKNNREENKVALFNKNRAENTLKIEKYDNLL	128
		::: ::: ::: ::: ::: ::: :::	:
QY	62	VKGGIASD---LIEIIGTSYTHAILPLLPSR--VEAOVQRDREVKBELFELSPKGFWLP	116
		::: ::: ::: ::: ::: ::: :::	
Db	129	INAFKRYDLRGCEVIITCAATHALLPINLRQAQKIATGVQSYYNTMGHEPNGIWLP	188
		::: ::: ::: ::: ::: ::: :::	
QY	117	EIAYDDIPTAILKDNGVEYIFADGEAMLFSAHUNSAIKPIKLPHLIKAAQREKRFRYS	176
		::: ::: ::: ::: ::: ::: :::	
Db	189	ECAYTYGIDNLSEFGIKYFISEKAIDVASP-----KPMY-----	224
		::: ::: ::: ::: ::: ::: :::	
QY	177	YLGLRELNRKAIKLFEGKVTLKAVKDIEAPVVAVNTAVMLGIRLPLMNPKVASMI	236
		::: ::: ::: ::: ::: ::: :::	
Db	225	-----GTNI-----PTAAPSGVCAGF	240
		::: ::: ::: ::: ::: ::: :::	
QY	237	EDKDNILLYGTDIFIG-----YRDIAGYRNSVEGLEVIDENSELCLPSELKH	286
		::: ::: ::: ::: ::: ::: :::	
Db	241	RMDS--SYQWNSDFMGPYGDFNVREPYRI-GFELPMYEIKPYINE--NGRIDTGFYK	295
		::: ::: ::: ::: ::: ::: :::	
QY	287	-----SGRE-LYLFRTSS---WAPDKSLRWREDGN-ARLNM-----	318
		::: ::: ::: ::: ::: ::: :::	
Db	296	YKITGSNGEKGIYNRENAMKWVESHASFCHDHQIDNAANMKDPPIITCPYDTELYG	355
		::: ::: ::: ::: ::: ::: :::	
QY	319	-----LSYNREGELALLAENSARGWEPLPERRLDAFRALNDWM	357
		::: ::: ::: ::: ::: ::: :::	
Db	356	HWWFEGPDPIAFIRKSADWTSYELITPTLEYLKNMWQCSSFP-----SSW	404
		::: ::: ::: ::: ::: ::: :::	
QY	358	RGENGE 363	
		:	
Db	405	-GENGD 409	

RESULT	6
Q973T0	PRELIMINARY;
ID	Q973T0
AC	Q973T0
DT	01-DEC-2001 (TREMBLrel. 19, Created)
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	HYPOTHEICAL PROTEIN ST0817.
GN	ST0817.
OS	Sulfolobus tokodaii.
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX	NCBL_TaxID=111955;

```

[1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RR PubMed-11572479;
RA Kavarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RA "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
RR EMBL; AP000983; BAB65830.1; -.
RW Hypothetical protein; Complete proteome.
RS SEQUENCE 443 AA; 53218 MW; 497087C264BC03C4 CRC64;

Query Match 7.6%; Score 142.5; DB 17; Length 443;
Best Local Similarity 23.0%; Pred. NC. 0.0024;
Matches 79; Conservative 41; Mismatches 134; Indels 89; Gaps 15;

QY 22 KVIEKAYIEVITLKEIPFGLNITGYTLKFL-----PKDIIDLVKGGIA 67
DB 46 RVNKKCIYIPATRIILL-EEIENG-ESEGYDFFFFSLGTFIEQAKWGKVDLFLFOOLSY 103
QY 68 SDLTIEIGSYTHAILPLL-PLSRVEAQVQDRVKVELFELSPKGFWLPFLAYDPIIPA 126
DB 104 THKVEFLSTYTHSITSLWEDLTENKEQVQMHKNLIRDYFQGPVTFENTELLTPRIYK 163
QY 127 ILKDNQYKFLADGGEAMFLFSAHNSAIKPIKPLPHILKIAQKRKRPYSYLLGLRELK 186
DB 164 EIEKLGFKVITTEGK-----ESLLKQSP-----NRVYRINDTKSLLLRNRL 207
QY 187 AIKLVEGKVTLLKAVKDIEAPVWVAVTAVMLGIGRLPLMNPKKVASWTEDKDNILYG 246
DB 208 SDIIAR-----FSNPKW-----DQVPL-TASKADWVSWSEGEI--- 241
QY 247 TDTEFTGYRDIAGYRMSVGLLEVIDELNSLCPLSELKHS-----RELYR----- 294
DB 242 -GLIFVDYETFGHEHWPESIGLDLF-----RWLPRELHRRGVEFKLPREVYKEYDEIV 294
QY 295 ---TSSWA---PKDS-----LRWREDEGNARLNMLSYNMRGE 326
DB 295 IDSTVSWADINKDESNLGNIMQWYDEWVRTEMLAKEGGE 337

RESULT 7
Q972NO PRELIMINARY; PRT; 895 AA.
ID Q972NO
AC Q972NO;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEITICAL PROTEIN ST1102.
GN ST1102.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
NCBI_TaxID=111955;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RR PubMed-11572479;
RA Kavarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RA "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
RR EMBL; AP000983; BAB65830.1; -.
RW Hypothetical protein; Complete proteome.
RS SEQUENCE 443 AA; 53218 MW; 497087C264BC03C4 CRC64;

```

```

QY 14 EIPKSEIPKVIKAYIPVIEL-----KKEIPFGLNITGYTLKPLPKDITDL---VK 63
DB 71 EYIKREFEKYMERKLSMEEDLERFKDEKLREAINF---MIGYF-----KDYYSYKSID 122
QY 64 GGIASDL-----IEIGTSYTHAILPLPLSR-VEAQQVQDRVKEBELFELSPKGF 114
DB 123 GNILKFELODEGVVEVITSAATHGYLPLGRDEATEAALLNGIKYKFKRGRGIW 182
QY 115 LPELAYDFI-----IPAILKONGYELFADGAMLFSAHLSAIPKIPK 158
DB 183 LPECAIRPDGLWSPSTGEVKKWRGIEHFLKFGIEYFFVE-----SHLIDK-GPVSL 234
QY 159 LYPHILIKAKRERERYISYLLGLRELKKAIKLVFEGKVTLKAVKADIEAVPVWA 212
DB 235 RYGNILPA-KIKRSTLRPF-----LKNIAVAFARNRET-----GIQVNSA 274

RESULT 9
Q9KD04 PRELIMINARY; PRT; 923 AA.
AC Q9KD04;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE BH1415 PROTEIN.
GN BH1415.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.,
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001512; BAB05134.1;
DR InterPro; IPR003803; DUF200.
DR InterPro; IPR001296; Glycos_transf_1.
DR Pfam; PF02651; DUF200; 1.
DR Pfam; PF00534; Glycos_transf_1; 1.
KW Complete proteome.
SQ SEQUENCE 923 AA; 106638 MW; 8C7B7D64F3603553 CRC64;

Query Match 6.8%; Score 128; DB 16; Length 923;
Best Local Similarity 20.1%; Pred. No. 0.11;
Matches 97; Conservative 60; Mismatches 167; Indels 158; Gaps 24;

QY 3 ALVFHNLQYAEIPKSE-----IPKVIKAYIPVIELIKEIPFGLNI----- 46
DB 7 SILVHAHLPPYVRHQEDRLERLWLFTEAMSETYIPLLWALEKLPVKHATISFTPPVMEML 66
QY 47 -----TGY-----TLKFLPKD-----IIDLVKG-----G 65
DB 67 SDPLVQTVRYNLHLENTEPQLKKEERTNDQNTQNLVQYKORYEKLKATFLQWDRNLLIG 126
QY 66 IASDLIE-----IIGTSYTHAILPLPLSR-VEAQQVQDRVKEBELFELSPKGFWLP 119
DB 127 FRS-LMENSQCILMTSAATHAFPPYKTKTEAIRAQVRHGIACFEQHFQKGLGFWLPECA 185
QY 120 YDPIIPAILKONGYELFADGAMLF---SAHLSAIPKIPLY-PH---LIKAQREKRF 172
DB 186 FSPGVDRILLFEIGIRYTFYDEGAVLTADPTPHKGS-----APIVSPHGIALFPRHTELSA 241
QY 173 RYTSYLLGLRELKKAIKLVFEGKVTLK-KVDIEAVPVK-----VAVTAVMLG 220
DB 242 KWNSSILG-----YPGDVYREYRIADREWDYIKPHVHKDGRIDTG---LK 288

```

```

QY 221 IGRPLMNPKK---VASWIEDK-----DNILLYGTDIIEFI 252
DB 289 YHRTIGHTHEESKDLVREWAERKVOEHANHTIGAIHHEIDQHGQNPFPYVMVTPFDALF 348
QY 253 GYRDITAGYRMSVEGLEEV-----IDELNSELCLPSELKHSGRELYLRTSSWAPDKSLRIWR 308
DB 349 GHWFEGPEW-IEALYEQGADRVSFITPELYLQRIHQ-DFQTAHVSFSTWGRDGYGHVWL 406
QY 309 EDEGNARLMLSYNWRGELALL-----AENSARQWEPLPERLDFAFRIN 355
DB 407 NDH-NAMRYRHYHREKDLAKIVAMYPOPTVLEKQAIQMVREW-----MLAVSS 455
QY 356 DW 357
DB 456 DW 457

RESULT 10
Q9V294 PRELIMINARY; PRT; 1362 AA.
AC Q9V294;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE AMYLOPULLULANASE.
GN APU OR PAS0122.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49104.1;
DR InterPro; IPR004300; Glyco_hydro_57.
DR Pfam; PF03069; Glyco_hydro_57; 1.
KW Complete proteome.
SQ SEQUENCE 1362 AA; 154578 MW; 277AFAB4E14860D1 CRC64;

Query Match 6.6%; Score 124; DB 17; Length 1362;
Best Local Similarity 21.5%; Pred. No. 0.39;
Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLKEEIPFGLNITGYTLKPLPKDITDLVYKGGIASDLIEIIGTSTHAILPLP--- 87
DB 236 VETVLKHQW-WLLNHT-----FEEHEKINLLG---NGNVEVTVVPTHTPIGLNDFGW 286
QY 88 LSRVQVQVORDREVKEELF---ELSPKGFWLPDELAYDPIIPAILKONGYELFADG---E 141
DB 287 YEDFQVQVKKANELYKELGACKVTPKGVAAESALNDKTLLEILAENGKWKWMTDQLVLE 346
QY 142 AMLFSAHLSAIPK-----IKPLYPHILKA 166
DB 347 KLGVPKTISSYKPVWAQFGDKKIYLFPRNHDLSDRVGFYAGMNQIDAVKNFVEELKI 406
QY 167 QREKPRYSYLLGLRELKKAIKLVFEGKVTLKAV-KDIEAVPVVAVNTAV-----ML 219
DB 407 QKONVDSGLVYVITLDGENPWEHYDFDKLFLEELYRLEELQKGLIRVTPSEYIEMF 466
QY 220 G---IGRLPLMNPKKVASWIEDKDNILLYGTDIETIGYRDIAGYRMSVEGLEVIDELNSE 277
DB 467 GDKANKLTKPMKRLDFTEDNVNALLKAKTL-----GELYDMVGVE----- 509
QY 278 LCLPSELKHSGRELYLRTSSWAPDKSLRIWREDEGNARLMLSYNMRGELALLAENS-D-A 336
DB 510 -----EMQWPESSWI-DGTLSTW---IGEPQENIAWVLYLARKALFENKDNV 553
QY 337 RGWEPLPERLDFAFRINDW 357

```

Db 554 KWNKAYEY---LFRAGSDW 571

RESULT 11

Q9HL91 PRELIMINARY; PRT; 357 AA.

AC Q9HL91

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE ALPHA-AMYLASE RELATED PROTEIN.

GN TA0339.

OS Thermoplasma acidophilum.

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;

OC Thermoplasma.

OX NCBI_TaxID=2303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;

RX MEDLINE=20479972; PubMed=11029001;

RA Ruepp A., Granl W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.,

RA "The genome sequence of the thermoacidophilic scavenger Thermoplasma

RT acidophilum";

RL Nature 407:508-513(2000).

DR EMBL; AL445064; CAC11483.1; -

KW Complete proteome.

SQ SEQUENCE 357 AA; 41845 MW; B8C1A97F4EB237F4 CRC64;

Query Match 6.5%; Score 122.5; DB 17; Length 357;

Best Local Similarity 16.6%; Pred. No. 0.075;

Matches 51; Conservative 59; Mismatches 126; Indels 71; Gaps 8;

QY 22 KVTEKAYIPVETLKEEIPFGNITGYL-----KFLPKDIIDLKGGIADSLIEIIGTS 77

Db 20 RIANNYPATQSLMDYGRSELSGTWEQSKYCPK-VIDVDIVDRSQOCELLSET 78

QY 78 YTHAIDPLPLSRVEAQVORDREVKEELFELSPKGFVLPPELAYDPIIPAILKONGYEYLF 137

Db 79 YHSLASINWDEEFVQVRMOESAIAKKTENYFVSRNTELIYNHDIHAEVAKGMGFRNII 138

QY 138 ADGEAMLFSAHLSAIPKIPLYPLHKAQREKRPY-----ISVLLGLRELRLKAIKLVF 192

Db 139 AEGTDDIASRY-----DVNRYAAPSGINLYLRNPLSDDISRF 178

QY 193 EGVTLKAVKDIKAVPVWAVNTAVMLGIGRPLMNPKKVASWIEDKNILLYGTDIETI 252

Db 179 SNRWDYPLTADKFARWISASSGEIVNL-----FM 209

QY 253 GYRDIAGYRMSVGLLEVI-----DELNSLCLPSLKHSGREL--YLRTSSWA-P 300

Db 210 DYTFGEHQRPETGIFELRYLPMYFEENDVHTILVREAEARHRTDFISVSKTTSWADK 269

QY 301 DKSLRIW 307

Db 270 NRDLASW 276

RESULT 12

C30246 PRELIMINARY; PRT; 324 AA.

AC C30246

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE METHANOL DEHYDROGENASE REGULATORY PROTEIN (MOXR).

GN AF2425.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;

OC Archaeoglobus.

OX NCBI_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,

RA Richardson D.L., Kerlavage A.R., Graham M.H., Kyrpides N.C.,

RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,

RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,

RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon Archaeoglobus fulgidus";

RL Nature 390:364-370(1997).

DR EMBL; AF001108; AAB91247.1; -

DR TIGR; AF2425; -

DR InterPro; IPR000523; Mg_chelatase_ChII.

DR Pfam; PF01078; Mg_chelatase; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 324 AA; 36609 MW; B0F64965EFC82F1F CRC64;

Query Match 6.4%; Score 120.5; DB 17; Length 324;

Best Local Similarity 21.7%; Pred. No. 0.095;

Matches 71; Conservative 60; Mismatches 121; Indels 75; Gaps 14;

QY 1 LRALVPHGNLQVAEIP---KSEIPKVIKAYIPVETLKEEIPFGLNITGYTLKFLPKD 57

Db 35 LAALINGNILEDPGLCKTKLLAKVFARV-----IGADYRRVQFTDPLPSD 82

QY 58 IIDLVKGGI---ASDLIEII-GTSYTHAIL-----PLPLSRVEAQVORDREVKEEL 105

Db 83 II-----GVKINRGDRFEVKGPIETNVLVLADEINRSPKTKQAALLEAMEEKQITVEGT 137

QY 106 FELSPKGFVLPPELAYDPIIPAILKONGYEYLFADGEAMLFSAHLSAIPKIPLYPLHIIK 165

Db 138 FSUSMPFFVL--ATQNP-----EQGTYPPLPEAQMDRFLMRPQGPESIE 182

QY 166 AQREKFRYISY-----LLGLRELRLKAIKLVPEKGVTLKAVKDIKAVPVWV 211

Db 183 EEMEILRRIRISMKDKDPTEDVSPVSLTFRRIQDAEAVYVDKSIKYLVRVRA---T 239

QY 212 AVNTAVMLGI---GRPLMNPKKVASWIEDKNIL--LYGTDIEFGYRDIAGYRMSVE 265

Db 240 REHELVELGSSPRGGALLKLARALAVMDGRDFVDPDDVKRVAVAEALHRLVILFEYAVE 299

QY 266 GLL--EVIDELNSLCLPSLKHSGRE 290

Db 300 GLRAEEVVEILNSVRVP---KYEAGE 323

RESULT 13

Q9VOM7 PRELIMINARY; PRT; 602 AA.

AC Q9VOM7

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 70.2 KDA PROTEIN.

GN PAB1857.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

OX NCBI_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORSAY;

RA Helling R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome

RT structure and evolution";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; AJ248285; CAB49676.1; -.
DR InterPro; IPR001395; Aldo_ket_red.
DR InterPro; IPR003803; DUF200.
DR InterPro; IPR003583; HH1.1.
DR Pfam; PF02651; DUF200; 1.
DR SMART; SM00278; HH1; 1.
DR PROSITE; PS00062; ALDOXETO_REDUCTASE_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 602 AA; 70219 MW; 326B1198D54DD43E CRC64;

Query Match          6.3%; Score 119; DB 17; Length 602;
Best Local Similarity 26.8%; Pred. No. 0.31;
Matches 42; Conservative 20; Mismatches 36; Indels 4; Gaps 4;

QY 14 EIPKSEIPKVIKAYIPVIEFLIKEIPFGNITGYTLKLPK-----DII 59
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 EYKAEFNRMYERKI-----KLMKEDLKADGKLRLNAIEFMKVFEDVYVWSKINGDII 125

QY 60 DLVKGGIASDLIIIGTSYTHAILPLSLR-VEAQVQRDREVKEELFELSPKGFWLPEL 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 126 GFEQLQDEGEVEITSAATHGYLPLGLRDEADAQLTGIRVYEFKFKKPRGIWLPEC 185

QY 119 AYDPT-----IPAILKDNNGEYLFAD 139
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 186 AYRPDGLKWSPTGEIKWRKGIEHFLKKYGLRFFVE 222

RESULT 14
ID O83377 PRELIMINARY; PRT; 526 AA.
AC O83377;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TP0358.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RA MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001215; AAC65344.1; -.
DR TIGR; TP0358; -.
DR InterPro; IPR003803; DUF200.
DR Pfam; PF02651; DUF200; 1.
KW Complete proteome.
SQ SEQUENCE 526 AA; 59699 MW; D5F0C32CADB6C41A CRC64;

Query Match          5.9%; Score 111; DB 16; Length 526;
Best Local Similarity 19.9%; Pred. No. 1.1;
Matches 43; Conservative 31; Mismatches 72; Indels 70; Gaps 7;

QY 3 ALVFGHNGLOYAE-----IPKSEIPKVIKAYIPVI---ETLKEIPFGNLT----- 46
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 AFVLDGNLPFVRGAGASLLAESRFFLEISYTYLPLRLCETLERVRVPFNISIAIGPVL 69

QY 47 -----TGYTLKFL--PKDIIDLKVG 64
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 CEMLANRVLMRYRALDALIEFGGERAIRLNRSLQERVQAEAVLRSLRSHRYDFDCDG 129
```

```
QY 65 GIASDL-----IEIGTSYTHAILPL---LPLSRVEAQVQRDREVKEELFELSPKGF 113
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 130 ALLERINEFFRTSGIELLATAVNCFLPFYQDPES-ISAQIMGLINKRKHFSSTPRGF 188

QY 114 WLPELAYDPIIPAILKDNNGEYLFADGEAMLFSAHL 149
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 189 YLPELGAPALERTIKSYGFSYILETHSLFLGTRV 224

RESULT 15
QY 972D2 PRELIMINARY; PRT; 447 AA.
AC Q972D2;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE ALPHA AMYLASE (EC 3.2.1.1).
GN SSO0988.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RC MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moers A., Erauso G., Fletcher C., Gordon P.M.K., Medina N., Peng X.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Theriault C., Tolstrup N.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006718; AA41260.1; -.
KW Hydrolyase; Glycosidase; Complete proteome.
SQ SEQUENCE 447 AA; 53641 MW; E3A45F31AC2D20EE CRC64;
```

```
Query Match          5.8%; Score 109.5; DB 17; Length 447;
Best Local Similarity 20.8%; Pred. No. 1.2;
Matches 72; Conservative 54; Mismatches 135; Indels 85; Gaps 17;

QY 17 KSEIPKVIKAYIPVIEFLI-----KEE---IPFGNITGYTLKFLP---KDIIDLKVG 64
   | : | : | | | : | : | : | : | : | : | : | : | : | : |
Db 41 KEIFERIKKNCYIPATNIIILSSIERAEENNNKYFFSISGTFLEQAERWGREVIELFQ 100

QY 65 GIASDLIEIGTSYTHAILPLL-PLSRVEAQVQRDREVKEELFELSPKGFWLPELAYDPI 123
   | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 101 LAYTHKVEFLAQTYHVSYLSWEDKSEWKEQVKMKDKTIKSYFGQYPTTFTEL----- 155

QY 124 IPAILKDNNGEYLFADGEAMLFSAHLSAIPKIPLYPHLIKAQREKRFYISYILGLRE 183
   | | | : | : | : | : | : | : | : | : | : | : | : |
Db 156 ---ITKDIVEEVEKMGFKMMLSEGTNRNLNGRSPNVYKLG-HEIRMLFRNYTL--- 208

QY 184 LRKAIKLVFEKGVTLKAVKDIE--AVPVVAVNTAVMLGIGRLPLMNPKKVASWTEKD 240
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 -----DDIAFFSNPNW-----DQYPL-TASKYADWLSRSE 238

QY 241 NILLYGTDTIEGYRDIAGYRMSVEGLEVID----ELNS---ELCLPSELKSHGRELY- 292
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 239 G----NVGLIFVDYETFGHEHREOTGILEFLKWLPIELNSKGVEMMPKEVYN---DVID 291

QY 293 ----LRTSSWA----PKDS----LRINREDEGNARLNMLSYNMRGE 326
   | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 EIEIAHTSWADIEKDEKSWLGNMQWYADDAVRAEMPSRELGNE 337
```

Search completed: June 2, 2002, 18:09:36
Job time: 218 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:01:42 ; Search time 32.53 Seconds
(without alignments)
1242.879 Million cell updates/sec

Title: US-09-886-400-4
Perfect score: 1877
Sequence: 1 LRLVFNGLQYAEIPKSEI.....RRLDAFRALYNDWRGENCEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1873	99.8	364	AAW34643	Thermotable alpha
2	145.5	7.8	655	AAW34643	Putative P. abyssi
3	143.5	7.6	649	AAW34643	Pyrococcus furiosus
4	124	6.6	1362	AAW34643	Putative P. abyssi
5	121.5	6.5	653	AAW34643	Super heat resista
6	104.5	5.6	360	AAW34643	S. epidermidis ope
7	104.5	5.6	376	AAW34643	S. epidermidis ope
8	104	5.5	329	AAW34643	Heat resistant mal
9	104	5.5	571	AAW34643	Arabidopsis thalia
10	104	5.5	571	AAW34643	Arabidopsis thalia
11	104	5.5	704	AAW34643	Arabidopsis thalia

12	104	5.5	704	21	AAW34643	Arabidopsis thalia
13	104	5.5	726	21	AAW34643	Arabidopsis thalia
14	104	5.5	726	21	AAW34643	Arabidopsis thalia
15	103.5	5.5	936	22	ABG18270	Novel human diagno
16	97	5.2	929	22	ABW70857	Drosophila melanog
17	96	5.1	1073	22	AAU36768	Staphylococcus aur
18	96	5.1	1147	22	AAU33972	Staphylococcus aur
19	93.5	5.0	317	22	AAW34643	Putative P. abyssi
20	92.5	4.9	480	21	AAW34643	Pinus radiata UGP
21	92.5	4.9	1668	22	AAU34158	Staphylococcus aur
22	92.5	4.9	2397	22	AAU36672	Staphylococcus aur
23	92	4.9	216	21	AAW34643	Arabidopsis thalia
24	92	4.9	216	21	AAW34643	Arabidopsis thalia
25	92	4.9	235	21	AAW34643	Arabidopsis thalia
26	92	4.9	1417	20	AAU31947	Plasmodium falcipa
27	91.5	4.9	473	21	AAW34643	A uridine diphosph
28	91.5	4.9	473	21	AAW34643	Zea mays protein f
29	91	4.8	364	21	AAW34643	Arabidopsis thalia
30	91	4.8	375	22	AAW34643	S. epidermidis ope
31	91	4.8	389	21	AAW34643	Arabidopsis thalia
32	91	4.8	852	22	AAW34643	Corynebacterium gl
33	91	4.8	852	22	AAW34643	C. glutamicum prote
34	91	4.8	4829	22	AAW34643	Human apoptosis-in
35	90.5	4.8	377	21	AAW34643	Zea mays protein f
36	90	4.8	844	21	AAW34643	Candida albicans C
37	90	4.8	1039	22	ABG15145	Novel human diagno
38	90	4.8	1248	22	ABW70857	Drosophila melanog
39	89.5	4.8	1057	19	AAW72067	HSV-2 strain SB5 C
40	89.5	4.8	1114	19	AAW72067	HSV-2 strain SB5 C
41	89.5	4.8	1114	19	AAW72067	HSV-2 strain SB5 C
42	89.5	4.8	3854	22	ABW70857	Drosophila melanog
43	89.5	4.8	5385	22	ABW70857	Drosophila melanog
44	89.5	4.8	5496	22	ABW70857	Drosophila melanog
45	89.5	4.8	8805	22	ABW70857	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW34643
ID AAW34643 standard; Protein; 364 AA.
XX AAW34643;
AC AAW34643;
DT 27-MAR-1998 (first entry)
DE Thermotable alpha-galactosidase AEDII12RA-alpha-gal-189C.
XX Alpha-galactosidase; alpha-galactosidase; thermostable enzyme;
XX Alpha-galactosidase; alpha-galactosidase; thermostable enzyme;
KW food processing; alpha glycoside hydrolysis; raffinose;
KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-189C.
XX
XX Thermococcus alcaliphilus strain AEDII12RA.
OS
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 329
FT /note= "encoded by CTT"
XX
XX W09732974-A1.
PD
PD 12-SEP-1997.
XX
XX 05-FEB-1997; 97WO-US01452.
XX
XX 08-MAR-1996; 96US-0613220.
XX
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
XX Murphy D, Reid J, Rudolph MJ;
XX
XX WPI; 1997-470541/43.
DR N-PSDB; AAT93753.
DR

XX Nucleic acid encoding alpha-galactosidase from Thermococcus
PT alcaliphilus - used in food processing to hydrolyse
PT alpha-galactosides, e.g. raffinose
XX
PS Claim 1; Fig 1; 32pp; English.
XX
CC This protein comprises AEDII12RA-alpha-gal-18GC, a claimed
CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
CC AEDII12RA, a bacterium that shows optimum growth at 85 deg C and
CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
CC AAR93753) encoding the alpha-galactosidase; (2) a vector containing
CC the polynucleotide or homologous or complementary sequences; (2)
CC host cells containing the vector; (3) a process for producing the
CC alpha-galactosidase in transformed or transfected host cells; an
CC enzyme showing at least 70% identity to alpha-galactosidase and
CC comprising at least 30 amino acid residues of its sequence; and (4)
CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
CC in sugar beet processing (raffinose inhibits crystallisation of
CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
CC and verbascone in beans and other gassy foods.
XX
SQ Sequence 364 AA;

Query Match 99.8%; Score 1873; DB 18; Length 364;
Best Local Similarity 99.7%; Pred. No. 1.8e-182;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LRALVFGNQLQYAEIPKSEIPKVIKAYIPVETLKKEELPFGCLNTGYTLKFLPKDIID 60
DB 1 LRALVFGNQLQYAEIPKSEIPKVIKAYIPVETLKKEELPFGCLNTGYTLKFLPKDIID 60
QY 61 LVKGGIASDLTEIIGTSTYTHAILPLLSRVEAQVORDREVKEELFSPKGFWLPDELAY 120
DB 61 LVKGGIASDLTEIIGTSTYTHAILPLLSRVEAQVORDREVKEELFSPKGFWLPDELAY 120
QY 121 DPTIPAILKONGYLYLFDAGEMLFSAHLNSAIKPKIPLYPHILKAQREKRYISYLLG 180
DB 121 DPTIPAILKONGYLYLFDAGEMLFSAHLNSAIKPKIPLYPHILKAQREKRYISYLLG 180
QY 181 LRRLKATKLVFEKGKVTLKAVKDIKAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 240
DB 181 LRRLKATKLVFEKGKVTLKAVKDIKAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 240
QY 241 NILLYGTDIEFIGYRDYAGYRMSVEGLLEVIDENSELCLPSELKHSGRRELYLRTSSWAP 300
DB 241 NILLYGTDIEFIGYRDYAGYRMSVEGLLEVIDENSELCLPSELKHSGRRELYLRTSSWAP 300
QY 301 DKSRLRWDEGNSARLNMLSYNGMGEGLALLAENSARGWEPPLPERRIDAFRAYINDWRGE 360
DB 301 DKSRLRWDEGNSARLNMLSYNGMGEGLALLAENSARGWEPPLPERRIDAFRAYINDWRGE 360
QY 361 NGEF 364
DB 361 ngef 364

RESULT 2
AAB96088
ID AAB96088 standard; Protein; 655 AA.
XX
AC AAB96088;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi alpha-amylase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX

PN FR2792651-A1.
XX
XX 27-OCT-2000.
PD
XX 21-APR-1999; 99FR-0005034.
PF
XX 21-APR-1999; 99FR-0005034.
PR
XX (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
DR
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
PS Claim 7; Pages 711-712; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAR86431 and AAR41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAR75903-AAR75920 and AAG66436.
XX
SQ Sequence 655 AA;

Query Match 7.8%; Score 145.5; DB 22; Length 655;
Best Local Similarity 22.4%; Pred. No. 1.1e-05;
Matches 90; Conservative 55; Mismatches 127; Indels 129; Gaps 22;
QY 25 EKAYIPVETLKKEELP---FGLNITGYTLKFLPK---DIIDLKGGIASDLTEIIGTSY 78
DB 28 ekayrpfleil--eeypnmkvaihsgilvewleenkpyidllkslvrkqgveivvvgf 85
QY 79 THAILPLLSRVEAQVORDREVKEELFSPKGFWLPDELAYDPTIPAILKONGYLYLFA 138
DB 86 yepvlaaipkedrleqlyllkewakki-gydakglwtervqgelvktlreagleyvvv 144
QY 139 DGEAMLFSAHLNSA-IKPKIPLYPHILKAQRE-----KRPY-----ISYLL 179
DB 145 d-----dyhfmaglskdqlfwpyytedggevityfp.deklrylipfrpdkvisylh 198
QY 180 GLRELKAIKLVFEGKVTLKAVKDIKAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
DB 199 slasedeskvavfh-----ddgekfgiw-----pm-----tyewvyek 231
QY 240 D-----NILLYGTDIE-----FIGYRDYAGY-----RMSVEG 266
DB 232 gwlrefeffrvssdeainimlyseylqkfkpgkglvlypiasyfemsewslpaggaklfve- 290
QY 267 LLEVIDENSELCLPSELKHSGRRELYLRTSSWAPDKSLRWDEGNSARLNMLSYNGM 323
DB 291 fveixkeln-----mferyrvirvggiw---Knf-fykypeanymhkmrlmls--- 334
QY 324 RGEALLAENSARGWEPPLPERRIDAFRAYIND--WRGENG 362
DB 335 ---rlirdpsarrf-----vraqcndaywhgvfg 362

RESULT 3
AAR47504
ID AAR47504 standard; Protein; 649 AA.
XX
AC AAR47504;

```

XX 07-JUL-1994 (first entry)
XX Pyrococcus furiosus alpha amylase.
XX
XX Pyrococcus furiosus; alpha amylase; liquefaction; polymers;
KW glucopolymers; thermostable.
XX
XX Pyrococcus furiosus.
XX
XX EP577257-A.
XX
XX 05-JAN-1994.
XX
XX 17-MAY-1993; 93EP-0303801.
XX
XX 09-JUN-1992; 92US-0893928.
XX
XX (UYJO ) UNIV JOHNS HOPKINS.
XX
XX Anfinsen CB, Laderman K;
XX
XX WPI; 1994-009532/02.
XX
XX Purified Pyrococcus furiosus alpha-amylase - used for the
XX industrial liquefaction of gluco-polymers at high temps.
XX
XX Claim 2; Figure 9; 4lpp; English.
XX
XX The purified Pyrococcus furiosus alpha amylase can act on substrates
CC with a low degree of polymerisation. e.g. glucose polymers as short
CC as maltotriose. The enzyme can be used for efficient industrial
CC liquefaction of glucopolymers at high temperatures.
XX
XX Sequence 649 AA;
SQ

Query Match 7.6%; Score 143.5; DB 15; Length 649;
Best Local Similarity 22.8%; Pred. No. 1.8e-05;
Matches 91; Conservative 59; Mismatches 118; Indels 131; Gaps 24;

QY 25 EKAYIPVETLKEEIP---FGLNITGYTLKPLPKDI---IDLKGGIASDLIEIGTS 77
DB 29 ekcywplfletl--eeypmkvahtsglliewl-qdirpeyidlrlslvkrqgvelvvag 85

QY 78 YTHAILPLPLSLRVEAQVORDEVKE--ELFELSPKGFWLPPELAYDPIIPAILKDNQY 135
DB 86 fypvlasip--kedrieqirlmkewaksigfdargvwltervwqpelvktlkesgidy 142

QY 136 LFDGGMFLPSAHLNS-----AIXPIKPLYPHLIKAOPEKRFYISVLL 179
DB 143 vlyvd-dyfhmsaelkeelywpyyrtedggevavipid-----ek-----Iryli 186

QY 180 GLRELKAKL---VPEGVTKAV--KDIEAVPVVWAVNTAVMLGIGRLPLMNPKKVAS 234
DB 187 pfprvdkvieylhslidgeskvavfhdggekfgiwpqiyewvy-----ekg 233

QY 235 WI-----EDKNILLYGTDIE-----FCGYRDIAGYRMSVEGLEVDLSELCL 280
DB 234 wlreffdrssdeklnlmlyteylekyxprgivyipiasy-----fem-----sewsl 281

QY 281 P-----SELKHS-----RELYLRTSSWAPDKSLRIWREDEGN---ARLNMLSYN 322
DB 282 pakqarfvefnelkvkfyekyrvfvggiw---knf-fykypesnymhkrmlmvskl 337

QY 323 MRGELALLAENSADARGWELPERRLDFAFIYND--WRG 359
DB 338 vrnn-----pearkyllraqcndaywhg 360

RESULT 4
AAB96091
ID AAB96091 standard; Protein; 1362 AA.

```

```

XX AAB96091;
XX
XX 29-OCT-2001 (first entry)
XX
XX Putative P. abyssi amylopullulanase.
XX
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX
XX Pyrococcus abyssi.
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX
XX Claim 7; Pages 715-719; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade.
XX
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX
XX AAH75903-AAH75920 and AAG66436.
XX
XX Sequence 1362 AA;
SQ

Query Match 6.6%; Score 124; DB 22; Length 1362;
Best Local Similarity 21.5%; Pred. No. 0.0056;
Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEEIPFGLNITGYTLKPLPKDIIDLKGGIASDLIEIGTSYTHAILPLP--- 87
DB 236 velvikhgm-wlnht-----feehkinlllg---ngnvevtvpythpigliindfgw 286

QY 88 LSRVEAQVORDEVKEELF---ELSPKGFWLPPELAYDPIIPAILKDNQYELFADG---E 141
DB 287 yedfdagvkkanelykeylgagkvtpkggwaesaalndktleilaengkwkwmtdqivle 346

QY 142 AMLFSAHLNSAKP-----IKPLYPHLIKA 166
DB 347 klgvptkiesytkpwwagfgdkkiylfprnhldsvrfgagmqdavknfveeliki 406

QY 167 QREKRPRTSYLILGLFELRKAIKLVFEGKVTLKAV-KDIEAVPVVWAVNTAV-----ML 219
DB 407 qknydgsivvyitidgenpwehyfpgklfleelyrleelqkkgllrtvtpseylemf 466

QY 220 G--IGRLPLMNPKKVASWIEDKDNILLYGTIDIEFGYRDIAGYRMSVEGLEVDLSE 277
DB 467 gdkankltpkmmkridfttednvnallkaktl-----jelydmvgvte----- 509

QY 278 LCLPSLKHSGRELYLRTSSWAPDKSLRIWREDEGNARLNMLSYNMRGELALLAENS 336
DB 510 -----emqwpesswi-dgtlstw-----igepgeniawwylyarkalfenkdnv 553

```

QY 337 RGWEPLERRLDARAIYNDK 357
 Db 554 kwnkayey---lfragsdw 571
 RESULT 5
 AAW54870
 ID AAW54870 standard; Protein; 653 AA.
 XX
 AC AAW54870;
 DT 01-SEP-1998 (first entry)
 DE Super heat resistant 4-alpha-glucanotransferase.
 KW Super heat-resistant 4-alpha-glucanotransferase; heat-treatment;
 KW alpha-1,4-glucan; alpha-1,4-glucoside bond.
 OS Pyrococcus sp.
 XX
 PN JP10150986-A.
 PD 09-JUN-1998.
 XX 21-NOV-1996; 96JP-0311117.
 XX 21-NOV-1996; 96JP-0311117.
 PA (BEAB-) BE ABLE KK.
 PA (NAGA-) NAGASE SEIRAGAKU KOGYO KK.
 DR WPI: 1998-375989/33.
 DR N-PSDB: RAV27026.
 XX
 PT New 4-alpha-glucanotransferase which has been heat-treated
 PT high temperatures to transfer at least one glucose unit
 XX
 PS Claim 4; Page 10-13; 18pp; Japanese.
 CC The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of
 CC 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at
 CC least 90% activity after heat-treatment at 100 deg. C for 30 minutes at
 CC pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan
 CC to alpha-1,4-glucan by an alpha-1,4-glucoside bond.
 XX
 SQ Sequence 653 AA;

Query Match 6.58; Score 121.5; DB 19; Length 653;
 Best Local Similarity 21.28; Pred. No. 0.0032;
 Matches 82; Conservative 64; Mismatches 141; Indels 99; Gaps 21;
 QY 25 EKAYIPVETTLKKEIP---PGLNITGYTLKFL---PKDIIDLKVGKIASDLIEIGTSV 78
 Db 28 ersyrfmetl--eeypmkvavhygspllewirdnkpehldllslvkrqgleivvagf 85
 QY 79 THAILPLLSRVEAQVQDRREVKEELFELSPKGFVLPPELAYDPIPAILKONGYEYLPFA 138
 Db 86 yepviasipkedrivqieklkefarnl-gyeargwiterwvqgelvksraagidyviv 144
 QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFERYISYLLGLRELKRAIK 189
 Db 145 d-----dyhfmagsqiskdelvpyytedggevltvfpideklr---ylipfrpvdktie 195
 QY 190 LVF-----EGKVTLKAV--KDIEAVPWPVAVNTAVMLGIGRLPLMNPKKVASWI----- 236
 Db 196 yhsiddgdeskvavfhddgekgfwpqtyewvy-----ekgwleffdrvs 242
 QY 237 -EDKNILLYGTDIE----FIGYRDIAGY-RMSVEGL-----LEVIDELNSELCLP 281
 Db 243 sderinlmlyseylrfrprgrglvlypiasyfemsewslparqaklfvfeveelk----- 296

QY 282 SELKHSGRELYLRTSSWAPDKSLRIWREDEGN---ARLNMLCYNNRGLALLAENSARG 338
 Db 297 kenkfdryrvfrggw---knf-fkypesnymhkrmlmvsavrnn----- 340
 QY 339 WEPLPERRLDARAIYND--WRGENG 362
 Db 341 ----pearefilraqcndaywhgvfg 362
 RESULT 6
 AAG82528
 ID AAG82528 standard; Protein; 360 AA.
 XX
 AC AAG82528;
 DT 03-SEP-2001 (first entry)
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
 KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
 KW vaccination; endocarditis.
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 PD 17-MAY-2001.
 XX 09-NOV-2000; 2000WO-US30782.
 XX 09-NOV-1999; 99US-0164258.
 PA (GLAX) GLAXO GROUP LTD.
 PA Kimmerly WJ;
 PI WPI: 2001-316495/33.
 DR N-PSDB: AAH53378.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis -
 XX
 PS Claim 18; Page 578-579; 2188pp; English.
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 360 AA;

Query Match 5.68; Score 104.5; DB 22; Length 360;
 Best Local Similarity 19.48; Pred. No. 0.069;
 Matches 58; Conservative 51; Mismatches 117; Indels 73; Gaps 12;
 QY 4 LVFHGNLQYAEIPKSEIPKVIKRAYIPVITLKEEIPFGLNITGYTLKPLPKDIDLK 63

```

Db 61 l1fkq-----vkrivedgysliirkliqnni--nlialhtnldvnpkgvnmrla 107
QY 64 GGIASDLIEIIGT--SYTHAILPLPLSRVEAQVORDREVKEELFELSPKGFWLPPELAYD 121
Db 108 dqglglenismitnsyykvtfipknyie-----dfkdslnel----- 147
QY 122 PIIPAILKONGYEYLFADGEAMLFSAHLNSAIKPKIPLYPHILKAQREKRPRIYSYLLGL 181
Db 148 ----glakegnyeycfeseg-----kgqfkpygdaspvlgklds-----leyv--- 187
QY 182 RELRKAIKLVF-----EGKVTLKAYKDIE--AVPV--WVAVNTAVMLGIGRLPLMNPVKV 232
Db 188 ----deiklefamkndeleitkrraildnhyetpvtfdikmkeseyglgiqngtmt 243
QY 233 ASWIEDKNILLYGTDIEFIGYRDIAGYRMSVEGLLEV-----IDELNSELCPLSELKH 286
Db 244 ldefseyakkqinipsvrytgqhdspikkkvaaligsigfeykasqigadvfvtgdikh 302

RESULT 7
AAG82506
ID AAG82506 standard; Protein; 376 AA.
XX
AC AAG82506;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2106.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI: 2001-316495/33.
XX
N-PSDB; AAH53356.
XX
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 569; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even

```

```

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 376 AA;
XX
XX Query Match 5.6%; Score 104.5; DB 22; Length 376;
XX Best Local Similarity 19.4%; Pred. No. 0.074;
XX Matches 58; Conservative 51; Mismatches 117; Indels 73; Gaps 12;
XX
QY 4 LVPHGMLQVAIEPKSEIPKVIKAYIPVIELTKEEIPFGNITGYTLKELPKDIIILVK 63
Db 47 l1fkq-----vkrivedgysliirkliqnni--nlialhtnldvnpkgvnmrla 93
QY 64 GGIASDLIEIIGT--SYTHAILPLPLSRVEAQVORDREVKEELFELSPKGFWLPPELAYD 121
Db 94 dqglglenismitnsyykvtfipknyie-----dfkdslnel----- 133
QY 122 PIIPAILKONGYEYLFADGEAMLFSAHLNSAIKPKIPLYPHILKAQREKRPRIYSYLLGL 181
Db 134 ----glakegnyeycfeseg-----kgqfkpygdaspvlgklds-----leyv--- 173
QY 182 RELRKAIKLVF-----EGKVTLKAYKDIE--AVPV--WVAVNTAVMLGIGRLPLMNPVKV 232
Db 174 ----deiklefamkndeleitkrraildnhyetpvtfdikmkeseyglgiqngtmt 229
QY 233 ASWIEDKNILLYGTDIEFIGYRDIAGYRMSVEGLLEV-----IDELNSELCPLSELKH 286
Db 230 ldefseyakkqinipsvrytgqhdspikkkvaaligsigfeykasqigadvfvtgdikh 288

RESULT 8
AAR94013
ID AAR94013 standard; Protein; 329 AA.
XX
AC AAR94013;
XX
DT 21-AUG-1996 (first entry)
XX
DE Heat resistant maleate dehydrogenase.
XX
KW Heat resistant maleate dehydrogenase; h-rMAD; NADH: L-aspartic acid;
KW alpha-ketoglutaric acid; glutamine oxalo-transaminase activity; GOT.
XX
OS Bacillus stearothermophilus ATCC 12016.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 185
XX FT /note= "Given in the specification as Var"
XX
XX JP08047389-A.
XX
XX 20-FEB-1996.
XX
XX 01-JUL-1994; 94JP-0151045.
XX
XX 03-JUN-1994; 94JP-0121629.
XX
XX 02-JUL-1993; 93JP-0164701.
XX
XX (TOYM ) TOYOBO KK.
XX
XX WPI: 1996-166248/17.
XX
XX N-PSDB; AAT17715.
XX
XX Protein having heat resistant maleate dehydrogenase activity - and
XX reagent contg. protein, NADH and L-aspartic and alpha-keto:glutaric
XX acid for determination of glutamine oxalo-transaminase activity
XX
XX Claim 8; Page 13-15; 17pp; Japanese.
XX
XX This sequence represents a protein having heat resistant maleate
XX dehydrogenase (h-rMAD) activity. The protein has a residual activity
XX after storage at 40 deg.C for 10 days of at least 60%, pref. 70%

```

[illegible]

```
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157755.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.5%; Score 104; DB 21; Length 571;
Best Local Similarity 22.5%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 112; Indels 122; Gaps 20;

QY 29 IPVETLIKEIPGLNITGYTLKFLPKDIDLVKGGIASDL-----IEIIGSYTHAIL 83
Db 101 ipile-mvpeelh----tvpyalsimgdpikmatlqldhqlptgvkiealrlltktml 156
QY 84 PLLPLSRVEAQVQRD-----RVKKELFEL-SPKGFWLPPELAYDP 122
Db 157 pl--lseggiipretllwlkllrscayansrihavqaeivlasgkdmmlpsqeeak 214
QY 123 IIPAIL-----KNGYEYLFADGEAMLFSAHLSAIPKIPKPLPHLIKAOREKFRY 174
Db 215 rihgilknscvrcfkdnghltlledsisll-----tvikgtg-----kyrrswrydl 261
QY 175 IS-----YLLG--LRELKKAIKLVF-----EGKVTLRKAVKDIEAVP---VWVAV 213
Db 262 vsdfilppskgelayaldevlgflrnavgsvffstmedgki----vkglagvdpkpgvllv 317
QY 214 NTAVMLGIGRLPLMNPKKVASWIEDKNILL-----YGTDFIEFIGY 254
Db 318 gyhlmgliepgmse-----afikek-nilfrgmahpvlvsdndpaxafdygdwkvfg- 370
QY 255 RDIAGYRMSVEGLLEVIDELNSLCLPS---ELKHSGRELYLRTSSWAPDKSLRIWREDE 311
Db 371 ----ayvtatnlfkildskshvllfpggarealhnrgey-----klixpeqq 415
QY 312 GNARL 316
Db 416 efvrn 420

RESULT 10
AAG48655
ID AAG48655 standard; Protein; 571 AA.
XX
AC AAG48655;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61465.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
```

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134321.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139219.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140951.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148241.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0156559.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	08-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	12-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159638.
PR	20-JUL-1999;	99US-0144884.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160767.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160770.
PR	21-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160814.
PR	22-JUL-1999;	99US-0145089.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145192.	PR	21-OCT-1999;	99US-0160815.
PR	22-JUL-1999;	99US-0145145.	PR	22-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	22-OCT-1999;	99US-0160981.
PR	23-JUL-1999;	99US-0145224.	PR	22-OCT-1999;	99US-0160989.
PR	26-JUL-1999;	99US-0145276.	PR	25-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	25-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	25-OCT-1999;	99US-0161406.
PR	27-JUL-1999;	99US-0145919.	PR	25-OCT-1999;	99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR	28-OCT-1999;	99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR	28-OCT-1999;	99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR	28-OCT-1999;	99US-0161920.
PR	02-AUG-1999;	99US-0146389.	PR	28-OCT-1999;	99US-0161920.
PR	03-AUG-1999;	99US-0147038.	PR	28-OCT-1999;	99US-0161992.
PR	04-AUG-1999;	99US-0147204.	PR	28-OCT-1999;	99US-0161993.
PR	04-AUG-1999;	99US-0147302.	PR	29-OCT-1999;	99US-0162142.
PR	05-AUG-1999;	99US-0147192.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			
PR	06-AUG-1999;	99US-0147416.			
PR	09-AUG-1999;	99US-0147493.			
PR	09-AUG-1999;	99US-0147935.			
PR	10-AUG-1999;	99US-0148171.			
PR	11-AUG-1999;	99US-0148319.			
PR	12-AUG-1999;	99US-0148341.			
PR	13-AUG-1999;	99US-0148565.			
PR	13-AUG-1999;	99US-0148684.			
PR	16-AUG-1999;	99US-0149368.			
PR	17-AUG-1999;	99US-0149175.			
PR	18-AUG-1999;	99US-0149426.			
PR	20-AUG-1999;	99US-0149723.			
PR	20-AUG-1999;	99US-0149929.			
PR	23-AUG-1999;	99US-0149802.			
PR	23-AUG-1999;	99US-0149930.			
PR	25-AUG-1999;	99US-0150566.			
PR	26-AUG-1999;	99US-0150884.			
PR	27-AUG-1999;	99US-0151065.			
PR	27-AUG-1999;	99US-0151066.			
PR	27-AUG-1999;	99US-0151080.			
PR	30-AUG-1999;	99US-0151303.			
PR	31-AUG-1999;	99US-0151438.			
PR	01-SEP-1999;	99US-0151930.			
PR	07-SEP-1999;	99US-0152363.			
PR	10-SEP-1999;	99US-0153070.			
<hr/>					
QY	29	IPVETLIKEIPFGLNITGYTLKFLPKDIIDLVRGGIASDL-----IEIIGTSYTHAIL 83	QY	29	IPVETLIKEIPFGLNITGYTLKFLPKDIIDLVRGGIASDL-----IEIIGTSYTHAIL 83
Db	234	lpile-mvpeelh---tvpalsfimgdpikmatigdnqlptgkviklqrktkml 289	Db	234	lpile-mvpeelh---tvpalsfimgdpikmatigdnqlptgkviklqrktkml 289
QY	84	PLPLSRVEXQVORD-----REVKEELFEL-SPKGFWLPCLAYDP 122	QY	84	PLPLSRVEXQVORD-----REVKEELFEL-SPKGFWLPCLAYDP 122
Db	290	pl--lseiglipretllwklkllrsgcayansrihavdaevlviagskdmmlpsqeaak 347	Db	290	pl--lseiglipretllwklkllrsgcayansrihavdaevlviagskdmmlpsqeaak 347
QY	123	IIPAIL-----KONGEYLFADGEAMLFSAHLSAIPKPKLYPHLIKAKQREKRPY 174	QY	123	IIPAIL-----KONGEYLFADGEAMLFSAHLSAIPKPKLYPHLIKAKQREKRPY 174
Db	348	rihgllkncsvrcfkdnghtlledsisll-----tvikgtg-----kyrrswrydl 394	Db	348	rihgllkncsvrcfkdnghtlledsisll-----tvikgtg-----kyrrswrydl 394
QY	175	IS-----YLLG--LRELRKAIKLVF-----EGKVTLKAVKDIENVP---VWVAV 213	QY	175	IS-----YLLG--LRELRKAIKLVF-----EGKVTLKAVKDIENVP---VWVAV 213
Db	395	vsdflpsskgelayaldevlgrnavgsvfstmedgkl-----vkglagvpdkgpvliv 450	Db	395	vsdflpsskgelayaldevlgrnavgsvfstmedgkl-----vkglagvpdkgpvliv 450
QY	214	NTAVMLGIGRLPLMNPKKVASWIEDKDNILL-----YGTDTIEFIGY 254	QY	214	NTAVMLGIGRLPLMNPKKVASWIEDKDNILL-----YGTDTIEFIGY 254
Db	451	gyhmlmglelcpmse-----afikek-nlfrgmahpviysndndpakafdygdwkvfg- 503	Db	451	gyhmlmglelcpmse-----afikek-nlfrgmahpviysndndpakafdygdwkvfg- 503
QY	255	ROIAGYRMSVEGLEVIDELNSELCLPS-----ELKHSRGRELYLFTSSWAPDKSLRWREDE 311	QY	255	ROIAGYRMSVEGLEVIDELNSELCLPS-----ELKHSRGRELYLFTSSWAPDKSLRWREDE 311
Db	504	----aypvtatnlfkldskshvllfggarealhurgecy'-----kllwpeqg 548	Db	504	----aypvtatnlfkldskshvllfggarealhurgecy'-----kllwpeqg 548
QY	312	GNARL 316	QY	312	GNARL 316

Query Match 5.5%; Score 104; DB 21; Length 704;

Best Local Similarity 22.5%; Pred. No. 0.221; Matches 82; Conservative 49; Mismatches 112; Indels 122; Gaps 20;

Db 549 efvm 553
|:
RESULT 12
AAG48654
ID AAG48654 standard; Protein; 704 AA.
XX
AC AAG48654;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136399.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140355.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141554.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 10-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

[illegible]

RESULT 14

AA48053
ID AAG48653 standard: protein: 726 AA.

AA
AC
AAG48653:

18-OCT-2000 (first entry)

xx
DE Arabidopsis thaliana protein fragment SEO ID NO: 61463.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX Arabidopsis thaliana.

XX
PN
EP1033405-A2XX
0006-685-90

XX
PF 25-FEB-2000: 2000EP-0301439

XX
25-FEB-1999. 0905-0121825.

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714

PR 16-APR-1999; 99US-0129845.
 PR 10-APR-1999; 99US-0130037.

PR 21-APR-1999; 99US-0130449.
PR 22-APR-1999; 99US-0130451.
PR 23-APR-1999; 99US-0130453.

